Helge Christian Flüß
Institut für Züchtungsforschung an landwirtschaftlichen Kulturen

Genetic analysis and construction of a genetic map of Russian dandelion (*Taraxacum kok-saghyz*) as an alternative resource crop for natural rubber

Dissertationen aus dem Julius Kühn-Institut

Julius Kühn-Institut
Bundesforschungsinstitut für Kulturpflanzen

Kontakt | Contact

Helge Christian Flüß
Julius Kühn-Institut
Bundesforschungsinstitut für Kulturpflanzen
Institut für Züchtungsforschung an landwirtschaftlichen Kulturen
Rudolf-Schick-Platz 3 a
18190 Sanitz, OT Groß Lüsewitz
Germany

Die Schriftenreihe „Dissertationen aus dem Julius Kühn-Institut" veröffentlicht Doktorarbeiten, die in enger Zusammenarbeit mit Universitäten an Instituten des Julius Kühn-Instituts entstanden sind.

The publication series „Dissertationen aus dem Julius Kühn-Institut" publishes doctoral dissertations originating from research doctorates and completed at the Julius Kühn-Institut (JKI) either in close collaboration with universities or as an outstanding independent work in the JKI research fields.

Der Vertrieb dieser Schriftenreihe erfolgt über den Buchhandel (Nachweis im Verzeichnis lieferbarer Bücher - VLB). Einige der Dissertationen erscheinen außerdem online open access und werden unter einer Creative Commons Namensnennung 4.0 Lizenz (CC-BY 4.0) zur Verfügung gestellt (https://creativecommons.org/licenses/by/4.0/deed.de). Die Schriftenreihe ist nachgewiesen in unserem Repositorium OpenAgrar: https://www.openagrar.de/receive/openagrar_mods_00005667.

The publication series is distributed through the book trade (listed in German Books in Print - VLB). Some of the dissertations are published online open access under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/deed.en). The publication series is documented within our repository OpenAgrar: https://www.openagrar.de/receive/openagrar_mods_00005667.

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie: detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Bibliographic information published by the Deutsche Nationalbibliothek
(German National Library)
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

ISBN 978-3-95547-159-0
ISSN (elektronisch) 2510-0602
ISSN (Druck) 2510-0599
DOI 10.5073/20251210-152438-0
D100

Herausgeber l Editor
Julius Kühn-Institut, Bundesforschungsinstitut für Kulturpflanzen, Quedlinburg, Deutschland
Julius Kühn-Institut, Federal Research Centre for Cultivated Plants, Quedlinburg, Germany

GENETIC ANALYSIS AND CONSTRUCTION OF A GENETIC MAP OF RUSSIAN DANDELION (*TARAXACUM KOK-SAGHYZ*) AS AN ALTERNATIVE RESOURCE CROP FOR NATURAL RUBBER

Justus Liebig University Giessen Institute of Plant Breeding and Agronomy I

Department of Plant Breeding Prof. Dr. Rod Snowdon

and

Julius Kühn Institute Research Centre for Cultivated Plants, Institute for Breeding Research on Agricultural Crops, Groß Lüsewitz

Dissertation

for a Doctorate Degree in Agricultural Sciences in the Faculty of Agricultural Sciences, Nutritional Sciences and Environmental Management

by

Helge Christian Flüß

Born in Nürnberg, Germany

Giessen, 2025

With permission of the Department of Agricultural Sciences,

Nutritional Science and Environmental Management,

Justus Liebig University Giessen

Examining Committee:

First Reviewer: Prof. Dr. Rod Snowdon

Second Reviewer: Prof. Dr. Benjamin Stich

Examiner: Prof. Dr. John Clifton-Brown

Examiner: Prof. Dr. Michael Frei

Chair of the Examining Committee: Prof. Dr. Jan Siemens

Date of Defence: 26.11.2025

For my son.

Acknowledgements

First, I would like to acknowledge funding for the projects TAKOWIND (FKZ: 22002312) and TAKOWIND II (FKZ: 22009216) from the German Federal Ministry of Food and Agriculture under coordination of the Federal Agency of Renewable Resources (FNR), Gülzow, Germany. In this context, I would like to thank Dr. Brigitte Ruge-Wehling, who gave me the opportunity to work on these projects.

My deepest thanks go to all the technical staff, who supported me throughout this thesis and beyond. First of all, Rico Fürstenberg, who introduced me to the lab and with his precise and diligent work, contributed significantly to these results. The same goes for Anja Katzel, Kathrin Schönherr and Beate Tessenow, who all cared cautiously for the plants, the lab and myself.

I would also like to thank all the project partners for sharing their passion for Russian dandelion. Especially Dr. Christian Schulze-Gronover, Dr. Nicole van Deenen and their team at the University of Münster, who helped a lot with the AFLP genotyping of the mapping population and with their scientific input. Also, many thanks to Katja Thiele for sharing her tremendous knowledge about this species and plant biology in general. Dr. Fred Eickmeyer of ESKUSA GmbH was always a huge inspiration with his innovative view on plant breeding and supported me greatly throughout the years.

Many thanks to Prof. Dr. Nils Stein, Dr. Axel Himmelbach and his technical crew at the IPK Gatersleben, who made it possible to conduct the Genotyping-by-Sequencing (GBS) project at their facilities. Next, I thank Dr. Jens Keilwagen and Thomas Berner at the JKI Quedlinburg, who introduced me to the JKI Galaxy platform and gave their input on the GBS data analysis.

I am very grateful for having Prof. Dr. Rod Snowdon as the first reviewer of this thesis. Thank you for never giving up on me and for all your support until the very end. The Justus-Liebig-University Giessen and the people at the department of plant breeding, especially Dr. Christian Obermeier, always had an open ear for me and helped me out a lot in the beginning of my scientific journey. Of course, I would also like to thank my old IFZ office buddy and long-term friend Prof. Dr. Kai Voss-Fels, for always keeping in touch, offering me advice or just having a laugh with me.

I would also like to express my gratitude to the new head of the JKI Institute for Breeding Research on Agricultural Crops, Prof. Dr. Benjamin Stich, for encouraging me to finish my thesis and offering me his guidance in this regard. Many thanks to all the new and old colleagues at the institute for all the scientific exchange and input.

Finally, yet importantly, my sincerest thanks go out to my closest friends and family, who supported me through all the years. My partner in crime and in love, Dr. Fluturë Novakazi, who is my biggest fan and critic at the same time, but is always there for me in good times and in bad

times. I could not be more grateful to have her in my life and without her, I probably could never have done all of this. I also thank my brothers Jan and Florian, who I can always count on when I need them. Most certainly, my biggest thanks go out to my mother, Elke. Her unconditional love, constant support and faith in me gave me the strength to finally finish this thesis.

There are probably still people that I forgot to mention. I humbly ask for their forgiveness and thank them nonetheless.

Content

Page

List of abbreviations III

List of figures IV

List of tables V

1 Introduction 1

1.1 Importance of natural rubber 1

1.2 *Taraxacum kok-saghyz* as an alternative resource for natural rubber 2

1.2.1 Historical context and strategic importance 2

1.2.2 Botanical characteristics and biology 3

1.2.3 Rubber biosynthesis and root physiology 4

1.2.4 Current research and breeding efforts 5

1.3 Molecular markers 6

1.3.1 Amplified Fragment Length Polymorphism (AFLP) markers 6

1.3.2 Simple Sequence Repeats (SSR) markers 7

1.3.3 Single Nucleotide Polymorphism (SNP) markers 9

1.4 Genetic linkage mapping and QTL analysis 10

1.4.1 Genetic linkage maps: Charting the code of inheritance 10

1.4.2 Quantitative Trait Loci (QTL): Unravelling complex traits 11

1.5 Objectives of the thesis 12

2 Material and Methods 13

2.1 Plant material 13

2.1.1 *Taraxacum* accessions and collections 13

2.1.2 F_1 mapping population 14

2.2 Phenotyping 14

2.2.1 *In vitro* propagation of the mapping population 14

2.2.2 Field trials and rubber quantification 14

2.2.3 Analysis of phenotypic data 15

2.3 DNA extraction and quantification 16

2.3.1 DNA extraction from plant material 16

2.4 Molecular marker analysis 17

2.4.1 Amplified Fragment Length Polymorphism (AFLP) analysis 17

2.5 Genotyping by sequencing (GBS) 21

2.5.1 Library construction and sequencing 21

2.5.2 NGS read mapping and SNP calling 22

2.6 Construction of genetic linkage maps 23
2.6.1 Marker scoring 23
2.6.2 Estimation of linkage groups and regression mapping 23
2.7 QTL analysis and mapping 24

3 Results 25
3.1 Assessment of genetic diversity 25
3.2 Phenotypic data of the mapping population 26
3.2.1 Rubber content and root weight 26
3.3 Genotyping-by-sequencing of the mapping population 28
3.4 Genetic maps of *Taraxacum kok-saghyz* 28
3.4.1 SNP calling and filtering 28
3.4.2 Construction of maternal and paternal genetic maps of TKS 29
3.4.2.1 Maternal map of high rubber producer E30-12, ... 29
3.4.2.2 Paternal map of low rubber producer E55-12 30
3.5 QTL analysis for rubber content 33

4 Discussion 37
4.1 Genetic diversity and population structure in *Taraxacum* 37
4.2 Phenotypic variability and heritability of rubber content 38
4.3 Construction of high-resolution genetic maps 39
4.3.1 Methodological challenges in Genotyping-by-Sequencing (GBS) 39
4.3.2 Marker integration and mapping efficiency 40
4.3.3 Quality and structure of the parental genetic maps 42
4.3.4 Map resolution and genome coverage 42
4.4 Identification and characterization of QTL for rubber content 45
4.4.1 Key QTL findings and the magnitude of their effects 45
4.4.2 Comparison with other studies and physical localization 45
4.4.3 Future directions and implications for breeding 47

5 Conclusions and outlook for future research 48

6 Summary 51

7 Zusammenfassung 52

8 List of Literature 53

Appendix I VI

Appendix II VII

Appendix III VIII

Declaration IX

List of abbreviations

AFLP	Amplified fragment length polymorphism
ANOVA	Analysis of variance
BLUEs	Best linear unbiased estimates
cM	centiMorgan
CIM	Composite interval mapping
CP	Cross-Pollination
CPTs	cis-prenyltransferases
CTAB	Cetyltrimethylammonium bromide
DMAPP	Dimethylallyl diphosphate
FAO	Food and Agriculture Organization of the United Nations
Gb	Gigabases
GBS	Genotyping by sequencing
GL	Groß Lüsewitz
GS	Genomic Selection
GWAS	Genome-Wide Association Studies
H^2	Broad sense heritability
IPP	Isopentenyl diphosphate
KASP	Kompetitive Allele Specific PCR
LG	Linkage group
LOD	Logarithm of odds
MAS	Marker-assisted selection
Mbp	Mega base pairs
MVA	Mevalonate
NGS	Next-Generation Sequencing
NMR	Nuclear Magnetic Resonance
NR	Natural rubber
PCR	Polymerase chain reaction
PMVK	Phosphomevalonate Kinase
PVE	Phenotypic variation explained
QLB	Quedlinburg
QTL	Quantitative Trait Loci
REFs	Rubber elongation factors
RFLP	Restriction fragment length polymorphism
SALB	South American Leaf Blight
SNPs	Single nucleotide polymorphisms
SR	Parkstetten (Straubing)
SRPPs	Small rubber particle proteins
SSR	Simple sequence repeat
TB	*Taraxacum brevicorniculatum*
TE	Tris-EDTA (buffer)
TKS	*Taraxacum kok-saghyz*
TO	*Taraxacum officinale*
USDA-ARS	United States Department of Agriculture - Agricultural Research Service
VCF	Variant Call Format

List of figures

Page

Figure 1.1: *Taraxacum kok-saghyz* plant in cultivation. Flower in full bloom is shown in the upper right corner (photo: H. Flüß). 4

Figure 3.1: Dendrogram illustrating the genetic relationships of *T. kok-saghyz* accessions (red, TKS USDA accessions are underlined), *T. officinale* (blue) and *T. brevicorniculatum* (green). 25

Figure 3.2: Density plot of rubber content in a mapping population of T. kok-saghyz (E30-12 x E55-12) over three years at the locations Groß Lüsewitz (GL), Quedlinburg (QLB) and Straubing (SR). Rubber content is shown as the percentage of root dry weight on the x-axis and the density is shown on the y-axis. 26

Figure 3.3: Boxplots showing the distribution of rubber content across three different locations Groß Lüsewitz (GL), Quedlinburg (QLB), Straubing (SR) over the years 2016 to 2018. Each boxplot represents the rubber content for a specific year, with the x-axis showing the years and the y-axis displaying the rubber content values in % of root dry weight. The boxes represent the interquartile range (IQR), the whiskers extend to the most extreme data points within 1.5 times the IQR, and the points outside the whiskers indicate outliers. 27

Figure 3.4: Maternal genetic linkage map of high rubber producer E30-12 based on 465 segregating molecular markers (SNP and AFLP) covering eight linkage groups. 31

Figure 3.5: Paternal genetic linkage map of low rubber producer E55-12 based on 576 segregating molecular markers (SNP and AFLP) covering eight linkage groups. 32

Figure 3.6: QTL analysis for rubber content on two linkage groups from the maternal map of parent E30-12. The plots display the LOD score (y-axis) at positions along the genetic map in centiMorgans (x-axis). The dashed line represents the significance threshold of LOD = 3.0. (a) shows a significant QTL peak (E30-12 QTL-1) on linkage group 2, and (b) shows a second significant QTL peak (E30-12 QTL-2) on linkage group 5. 33

Figure 3.7: QTL analysis for rubber content showing a significant QTL peak on linkage group 1 from the paternal map of parent E55-12. The plot displays the LOD score (y-axis) at positions along the genetic map in centiMorgans (x-axis). The dashed line represents the significance threshold of LOD = 3.0. 34

Figure 3.8: Physical maps of TKS showing the relative positions in Mbp (Mega base pairs) of genes involved in rubber biosynthesis in black colour, QTL identified by Yang et al. (2023) in pink colour, aligned GBS contigs of the QTL of this study which aligned to unique locations in red colour, contig c_55650 which aligned to three chromosomes in blue colour and contig c_50405 which aligned to ChrA1 and ChrA7 in orange colour. For MapChart data with the exact positions see Appendix III. 36

List of tables

Page

Table 2.1: List of *Taraxacum* accessions used for AFLP analysis with respective origin and coordinates (where available) as well as providing source. 13

Table 2.2: Distinct environmental conditions of the three field trial locations used for phenotyping the mapping population: Parkstetten (SR), Quedlinburg (QLB), and Groß Lüsewitz (GL). 15

Table 2.3: Reaction components for the double digestion of genomic DNA with *Eco*RI and *Mse*I restriction enzymes. 18

Table 2.4: Nucleotide sequences for the *Eco*RI and *Mse*I adapters that are subsequently ligated to the digested DNA. 18

Table 2.5: Components for ligation of adapters of one template. 18

Table 2.6: PCR components for pre-amplification of one sample for the AFLP marker analysis protocol. 19

Table 2.7: PCR program for pre-amplification for the AFLP marker analysis protocol. 19

Table 2.8: PCR components for selective amplification of one sample for the AFLP marker analysis protocol. 20

Table 2.9: PCR program for selective amplification for the AFLP marker analysis protocol. 20

Table 3.1: Two-factorial Anlaysis of variance (ANOVA) for rubber content in the mapping population E30-12 x E55-12. 27

Table 3.2: Summary statistics of the maternal genetic linkage map (E30-12). 30

Table 3.3: Summary statistics of the paternal genetic linkage map (E55-12). 30

Table 3.4: Marker intervals for E30 QTL-1, E30 QTL-2 and E55 QTL-3 on the corresponding linkage group (LG). The position of the individual SNP markers is given in centiMorgan. All SNP markers with a LOD > 3 and the explained phenotypic variance (PVE %.) are listed. GBS contig refers to the respective contig of the pseudo-reference sequence (107 bp length) on which the respective SNP marker was called. 34

Table 3.5: Blast results showing the positions of the respective QTL GBS contigs containing the significant SNP markers on the chromosomes of the reference genome. 35

Table 4.1: Comparison of genetic linkage maps for TKS 43

1 Introduction

Global demand for natural rubber continues to grow, yet the industry faces significant vulnerabilities (Warren-Thomas et al., 2015). The reliance on a single species, *Hevea brasiliensis*, leaves the supply chain susceptible to diseases and economic pressures, including rising labour costs and competition for land (Rocha et al., 2011). These factors highlight the urgent need to diversify the sources of natural rubber. Russian dandelion (*Taraxacum kok-saghyz*, TKS), an alternative that is suitable to grow in a temperate climate, has emerged as a promising candidate (Cornish, 2017). Historically important during wartime, renewed interest in this species has spurred modern research into its genetic characteristics and potential for commercial cultivation.

This thesis will address critical knowledge gaps in the genetic improvement of TKS. The research will utilize molecular markers to assess genetic diversity and construct high-resolution linkage maps. These foundational tools enable the mapping of quantitative trait loci (QTL) to pinpoint genomic regions associated with natural rubber content. By doing so, the study aims to provide crucial insights into the genetic architecture of this complex trait, paving the way for advanced breeding strategies like marker-assisted and genomic selection to accelerate the development of a resilient and sustainable rubber crop.

1.1 Importance of natural rubber

Natural rubber (NR) is a biopolymer with unique and irreplaceable physical and chemical properties, hence making it a material of great economic and strategic importance (Cherian et al., 2019). Composed primarily of cis-1,4-polyisoprene, NR is essential for the manufacture of over 50,000 different products, ranging from medical devices to tires for aircraft and heavy-duty trucks. The tire industry alone consumes over 75% of the total NR produced globally (Yang et al., 2023). Synthetic rubber (SR), derived from petroleum, has been used to supplement NR production and currently holds the larger share of the overall rubber market. However, SR's industrial performance is not comparable to that of NR in many high-performance applications like airplane tires for example (Cornish, 2017).

Global demand for NR is steadily increasing, driven by industrialization and the growing automotive and transport sectors in emerging economies (Warren-Thomas et al., 2015). According to the International Rubber Study Group (IRSG), global consumption was projected to reach 16.5 million metric tons by 2023 (Kuluev et al., 2023). A report from the Food and Agriculture Organization of the United Nations (FAO) indicates that global production of NR increased by 109% and harvest area by 72% between 2000 and 2020, with top-producing countries including Thailand, Indonesia, and Vietnam (Raleira & Drost, 2024).

For more than a century, the Pará rubber tree (*Hevea brasiliensis*) has been the exclusive commercially relevant source of NR, with over 90% of global production concentrated in tropical and subtropical regions of Southeast Asia (Luo et al., 2018). The rapid expansion of rubber tree plantations came at the expense of biodiversity rich rainforests and raises significant ecological concerns (Warren-Thomas et al., 2018). The reliance on a single species for NR production represents several critical vulnerabilities. For example, the crop's lack of genetic diversity makes it highly susceptible to catastrophic diseases such as the South American Leaf Blight (SALB), which has historically crippled production in Brazil (Guyot & Le Guen, 2018). Furthermore, the manual process of tapping latex is leading to rising costs and a shift in land use towards more profitable, less labor-intensive crops like palm oil.

Despite these downsides, NR from *H. brasiliensis* still remains the most significant rubber source, accounting for approximately 46% of the global rubber market (Adak et al., 2025).

The convergence of identified factors—growing global consumption, insecurity in the form of disease, competition for land, rising labour costs—should offer a compelling case for the necessity of diversifying the global natural rubber supply. The urgent need to develop domestically-grown rubber crop alternatives and the capacity for wider environmental suitability is also vital for the economic and sustainable security of the natural rubber industry.

While several plant species are considered alternatives for producing economically viable NR with high molecular weight, their climatic and agronomic requirements vary significantly (Mooibroek & Cornish, 2000). For instance, guayule (*Parthenium argentatum*) is adapted to the arid conditions of Mexico and the Southwestern U.S. (Cornish, 2001), and *Lactuca sativa* demands nutrient-rich soils for optimal growth (Bushman et al., 2006). In this context, Russian dandelion (TKS) emerges as one of the most promising candidates for implementation in temperate climates. Its primary advantages are a short life cycle and notable adaptability to a wide range of soil types, a trait reflective of its native habitat in Kazakhstan and China (Uteulin et al., 2023; Whaley & Bowen, 1947).

1.2 *Taraxacum kok-saghyz* as an alternative resource for natural rubber

1.2.1 Historical context and strategic importance

Taraxacum kok-saghyz (TKS) Rodin, commonly known as Russian dandelion or Kazakh dandelion, first emerged as a critical agricultural crop during World War II, when traditional rubber supply chains from Southeast Asia were severely disrupted. This made this unassuming perennial plant a vital strategic resource for the Soviet Union and other allied nations (Whaley & Bowen, 1947). The plant's ability to produce high-quality natural rubber in temperate climates made it an invaluable alternative to the tropical *H. brasiliensis*, which dominated global rubber production but was geographically inaccessible during wartime.

During the 1930s and 1940s, extensive cultivation programs were established across the Soviet Union, with thousands of hectares dedicated to rubber dandelion production. These early cultivation efforts demonstrated the plant's potential as a viable rubber crop, though post-war availability of Southeast Asian rubber led to the abandonment of most commercial TKS programs. While early breeding progress and germplasm collections at the Vavilov Institute were partially lost over time (Kirschner et al., 2013), challenges in NR production associated with *H. brasiliensis* prompted renewed interest in TKS in the early 2000s. The foundation for contemporary breeding programs was firmly established in 2008 with the collection of new TKS accessions in Kazakhstan (USDA-ARS accessions) (Hellier, 2011). This vital new genetic material has since catalysed numerous breeding programs worldwide, especially in the EU, and has been underpinned by large-scale, publicly funded European research initiatives like EU-PEARLS (GA No. 212872), DRIVE4EU (GA No. 613697), TARULIN (GA No. 0315971), TAKOWIND (GA No. 22002312), and EVITA (GA No. 031A285A).

1.2.2 Botanical characteristics and biology

TKS is a diploid (2n = 16) perennial herbaceous species belonging to the family Asteraceae with a genome size of around 1.29 Gigabases (Gb) (Lin et al., 2022). It can be distinguished from its close relatives by several key biological characteristics that make it particularly suitable for rubber production (Lin et al., 2018). Unlike the common dandelion (*Taraxacum officinale*, TO), which is typically triploid and reproduces asexually through apomixis, TKS exhibits obligate sexual reproduction due to its self-incompatibility system (Yuan et al., 2020). This reproductive strategy, while complicating breeding efforts, provides significant advantages for genetic improvement programs by maintaining genetic diversity and enabling controlled hybridization (Dijk et al., 1999).

The plant's morphology is characterized by a robust taproot system that serves as the primary site of rubber accumulation. The roots can extend 15-30 cm in length and develop a complex network of laticifers—specialized cells that synthesize and store natural rubber. These laticifers form an interconnected network throughout the root system, creating reservoirs of high-quality cis-1,4-polyisoprene that under optimal conditions can constitute 6-20 % of the root's dry weight (Cherian et al., 2019).

The above-ground portion of TKS consists of a basal rosette of deeply lobed leaves that emerge from a shortened stem. The plant produces characteristic yellow composite flowers typical of the Asteraceae family, borne on hollow scapes that can reach 20-40 cm in height (**Figure 1.1**). The flowering period typically occurs from late spring through early autumn, with individual plants capable of producing multiple flower heads throughout the growing season (Kreuzberger et al., 2016).

Figure 1.1: *Taraxacum kok-saghyz* plant in cultivation. Flower in full bloom is shown in the upper right corner (photo: H. Flüß).

1.2.3 Rubber biosynthesis and root physiology

The production of NR in TKS occurs through a complex biosynthetic pathway localized primarily in the root laticifer system. Unlike *H. brasiliensis*, where rubber is harvested from stem latex, TKS accumulates rubber in specialized root cells, requiring different harvesting and processing approaches. The overall NR quality of TKS, e.g. molecular weight, is comparable to *H. brailiensis*, with slight differences in lipid composition, but similar pathway of cis-1,4-polyisoprene synthesis (Bae et al., 2020). The rubber biosynthesis pathway begins with the mevalonate pathway, where acetyl-CoA is converted through a series of enzymatic reactions to produce isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), the fundamental building blocks of natural rubber (Cherian et al., 2019).

The polymerization of these isoprenoid units is catalysed by cis-prenyltransferases (CPTs), which represent a critical control point in rubber biosynthesis (Niephaus et al., 2019). Recent transcriptomic and proteomic analyses have identified multiple CPT isoforms in TKS, with tissue-specific expression patterns that correlate with rubber accumulation patterns (Zhang et

al., 2025). These enzymes work in conjunction with rubber transferase complexes to produce the high-molecular-weight cis-1,4-polyisoprene chains that constitute natural rubber.

The rubber particles in TKS roots are stabilized by a complex array of proteins, including small rubber particle proteins (SRPPs) and rubber elongation factors (REFs), which play crucial roles in maintaining particle integrity and regulating polymer chain length (He et al., 2024). These proteins form a protective monolayer around the rubber core, preventing particle aggregation and facilitating the controlled synthesis of uniform polymer chains. Understanding the function and regulation of these proteins has become a key focus of plant biotechnology efforts aimed at improving rubber yield and quality.

1.2.4 Current research and breeding efforts

Research and breeding programs worldwide are focusing on developing TKS cultivars with enhanced rubber content, improved agronomic characteristics, and better adaptation to local growing conditions. The Ohio State University has been particularly active in developing TKS as an alternative rubber crop, conducting extensive field trials and genetic studies (Cornish et al., 2016). German institutions, including the University of Münster, Fraunhofer Society and Julius Kühn-Institute, have made substantial contributions to understanding rubber biosynthesis, self-incompatibility, as well as agronomy and breeding research topics (Eggert et al., 2018; Kaiser et al., 2025; Wollenweber et al., 2021). The development of specialized cultivation and processing equipment designed specifically for root rubber production is progressing, with several prototype systems like the “Taraxagum Lab Anklam” by Continental (tire producer), showing commercial potential.

Despite significant progress in TKS research and breeding, several challenges remain before this species can become a commercially viable rubber source. The self-incompatibility system complicates seed production and cultivar development. Although rubber content and yield could be increased through recurrent selection (Hodgson-Kratky et al., 2017), it still lags behind those achievable with *H. brasiliensis* under optimal tropical conditions (2,200 to 2,400 kg NR per ha (Priyadarshan, 2017)). Recent studies have shown that the currently available TKS material has low to moderate genetic diversity, indicating a need for enhancing the gene pool of TKS (Kaiser et al., 2025; Nowicki et al., 2019).

Research has examined the potential for interspecific hybridization between TKS and related species, particularly TO because of its high vigour and biomass potential (Drummond & Vellend, 2012). Nevertheless, in order to make this potential accessible, more research is needed for overcoming crossing barriers, understanding recombination events and tools for targeted introgression of the desired traits.

Advances like the completion of a high-quality genome assembly of TKS by Lin et al. (2022) will enable the further development of molecular markers for breeding applications like marker assisted selection (MAS) based on identified QTL associated with rubber content for example (Yang et al., 2023).

Furthermore, genetic transformation technologies have opened new avenues for TKS improvement. Successful protocols for Agrobacterium-mediated transformation have been developed, enabling the introduction of genes that enhance rubber biosynthesis or improve plant performance under stress conditions (Peng et al., 2023). Gene editing technologies, particularly CRISPR/Cas systems, are being employed to modify key regulatory genes in the rubber biosynthesis pathway, with early results showing promising increases in rubber accumulation (Mahmoudieh et al., 2025).

1.3 Molecular markers

Historically, choosing individuals based on desired morphological and agronomic traits has been essential to the systematic improvement of crop plants (J. Liu et al., 2021). Phenotypic expression is often influenced by environmental factors or experimental error, an effect that may only become apparent at late developmental stages and makes selection only based on phenotype difficult and error-prone. A paradigm shift in plant breeding was brought about by the development of molecular genetics, which provided a set of instruments that could get around these restrictions. By making it possible to directly detect genetic polymorphism at the DNA level, the creation of molecular markers—recognizable DNA sequences linked to a particular locus on a chromosome—revolutionized the field (Kumar et al., 2024). This shift from phenotypic to genotypic selection provided breeders with a powerful, environment-independent tool to accelerate genetic gain.

1.3.1 Amplified Fragment Length Polymorphism (AFLP) markers

The development of Amplified Fragment Length Polymorphism (AFLP) markers represents a cornerstone technology for genomic analysis across diverse organisms. Originally developed by Vos et al. (1995), the AFLP methodology combines the reliability of restriction fragment length polymorphism (RFLP) with the power and flexibility of polymerase chain reaction (PCR) amplification, creating a highly reproducible fingerprinting method that requires no prior sequence knowledge (Vos et al., 1995; Vuylsteke, Peleman, & van Eijk, 2007). This technique has demonstrated exceptional utility in genetic diversity assessment, linkage mapping, and marker-assisted breeding programs, particularly for species with limited genomic resources.

A complex three-step procedure is used in the AFLP protocol to guarantee specificity and best reproducibility. Initially, total genomic DNA undergoes simultaneous digestion with two restriction enzymes, typically a rare-cutting enzyme (6-8 base recognition sequence) such as

EcoRI and a frequent-cutting enzyme (4-base recognition sequence) such as MseI or TaqI (Vos et al., 1995). This dual-enzyme approach generates restriction fragments of optimal size for subsequent PCR amplification while maintaining comprehensive genome coverage. Following digestion, double-stranded adapters complementary to the restriction site overhangs are ligated to the DNA fragments using T4 DNA ligase. A critical feature of this step is that the adapter sequences prevent restoration of the original restriction sites, enabling simultaneous restriction and ligation reactions in a single tube (Vuylsteke et al., 2007).

The selective amplification phase represents the core aspect of AFLP technology. PCR primers consist of three distinct regions: a core sequence complementary to the adapter, the restriction enzyme recognition sequence, and one to three selective nucleotides extending beyond the restriction site (Vos et al., 1995). Under stringent annealing conditions, only restriction fragments possessing nucleotides complementary to the selective extensions undergo amplification, effectively reducing the complexity of the resulting DNA fingerprint. This selective amplification mechanism enables researchers to adjust the number of amplified fragments by modifying the number of selective nucleotides, for each additional selective base reduces fragment complexity approximately four-fold (Mueller & Wolfenbarger, 1999)

Technological advances have further enhanced AFLP's utility and accessibility. The transition from radioactive labelling to fluorescent detection systems has eliminated safety concerns while improving data quality and throughput. Automated capillary electrophoresis systems, such as those manufactured by Applied Biosystems and Beckman-Coulter, enable high-resolution fragment separation and facilitate multiplexed reactions using different fluorescent dyes (Castiglioni et al., 1999). These developments have transformed AFLP from a labour-intensive technique requiring specialized facilities into a routine laboratory method suitable for large-scale genomic studies.

For emerging alternative crops such as TKS, AFLP markers represented a particularly valuable genetic analysis tool at the time of study. Despite being considered an older molecular marker technique, AFLP's independence from prior sequence information and comprehensive genome coverage make it ideally suited for initial genetic characterization of species lacking extensive genomic resources (Powell et al., 1996). Unlike most next-generation sequencing approaches that require reference genomes or comprehensive single nucleotide polymorphism (SNP) databases, AFLP fingerprinting can almost always be immediately applied to any organism regardless of the availability of genomic infrastructure.

1.3.2 Simple Sequence Repeats (SSR) markers

Simple Sequence Repeats (SSR), also known as microsatellites, represent one of the most valuable classes of molecular markers in modern plant breeding and genetic research. These

DNA sequences consist of tandemly repeated short nucleotide motifs, typically 1-6 base pairs in length, distributed throughout plant genomes at high frequencies (Vieira et al., 2016).

SSR markers are characterized by their hypervariable nature, arising from differences in the number of repeat units at specific loci between individuals (Vieira et al., 2016). These sequences are found abundantly across plant genomes, with frequencies depending on the species (Zhang et al., 2010). The basic methodology for SSR marker development involves several key steps: genome or transcriptome sequencing, identification of SSR-containing sequences using bioinformatics tools, primer design flanking the repeat regions, and validation through PCR amplification and polymorphism screening (Ou et al., 2025; Ouyang et al., 2018). Modern approaches utilize next-generation sequencing technologies to identify SSRs from whole genome sequences or transcriptome data, further streamlining SSR discovery and enabling high-throughput marker development (Yan et al., 2025).

The advantages of SSR markers compared to dominant or bi-allelic markers make them particularly valuable for plant breeding applications. Their codominant inheritance pattern allows discrimination between homozygous and heterozygous genotypes, providing complete genetic information at each locus (Ahmad et al., 2018). The multiallelic nature of SSRs generates high levels of polymorphism, with studies reporting 2-24 alleles per locus across different species (Alekseeva et al., 2023). Additionally, the reproducibility and transferability of SSR markers across related species represent significant advantages, with cross-species amplification success rates of 10 - 40 %, enabling comparative genetic studies and marker development in species lacking genomic resources (Ouyang et al., 2018)

Despite their advantages, SSR markers face several limitations that constrain their widespread adoption. The development process is time-consuming and costly, requiring substantial investment in sequencing, primer design, and validation (Ahmad et al., 2018). Technical challenges include the occurrence of null alleles due to primer binding site mutations, which can lead to allelic dropout and scoring errors (Hasnain & Mehvish, 2020). Additionally, the presence of stutter bands during PCR amplification can complicate allele scoring, particularly for dinucleotide repeats. The labour-intensive nature of electrophoresis-based detection methods and the requirement for specialized equipment and expertise further limit their accessibility in resource-constrained settings.

Nevertheless, SSR markers are still especially useful for the construction of linkage maps in outcrossing species or pseudo-testcross populations. Their multiallelic features allows for more than two different alleles, compared to bi-allelic SNPs for example, making them especially useful for map integration and use as anchor markers (Ahmad et al., 2018).

1.3.3 Single Nucleotide Polymorphism (SNP) markers

Single nucleotide polymorphisms (SNPs) are variations at a single base pair in the DNA sequence and represent the most abundant form of genetic variation in plant genomes (Morgil et al., 2020). Their primary advantage over older marker systems, such as SSRs or AFLPs, lies in their suitability for high-throughput, automated genotyping. This feature, combined with their widespread and uniform distribution across the genome, makes them highly cost-effective and efficient for large-scale genetic studies. The digital and robust nature of SNP data also ensures high reproducibility, allowing for reliable comparison of results across different laboratories (Yirgu et al., 2023).

The discovery and genotyping of SNPs have been revolutionized by Next-Generation Sequencing (NGS) technologies (Kilian et al., 2012). While early efforts involved mining existing sequence databases, modern approaches like whole-genome resequencing or Genotyping-by-Sequencing (GBS) can simultaneously identify and score hundreds of thousands of SNPs across large populations (Elshire et al., 2011; Voss-Fels & Snowdon, 2016).

High-throughput SNP arrays represent another powerful tool for application of SNP markers, offering standardized genotyping across large populations. The development of species-specific and even multi-species arrays, such as the 580k SNP Axiom Rice Genotyping Chip for rice (*Oryza sativa*) or the Infinium™ Wheat Barley 40K v1.0 BeadChip for wheat and barley (Keeble-Gagnère et al., 2021), enables rapid and cost-effective genotyping while maintaining high data quality. These arrays allow for the simultaneous genotyping of thousands of SNPs with well-characterized performance metrics and are therefore routine tools in many commercial breeding programs.

SNP arrays provide several advantages over traditional markers like microsatellites, including reduced time and cost, higher throughput, and greater reproducibility. However, array-based approaches require prior sequence information for probe design and may miss novel variants not represented on the array. The transferability of SNPs between different varieties within a species can also vary significantly, requiring validation when applying markers across diverse genetic backgrounds (Voss-Fels & Snowdon, 2016).

In breeding research, SNP markers have become indispensable tools and often the marker system of choice, for example when used to assess genetic diversity and population structure, which helps breeders understand the genetic relationships within their germplasm and select diverse parents for creating new crosses (Kumar et al., 2024). SNP markers are also used to construct the high-density genetic maps necessary for QTL mapping, a technique that identifies genomic regions controlling complex traits like yield or stress tolerance (Collard et al., 2005).

Furthermore, SNPs are the foundation for advanced breeding strategies. In Genome-Wide Association Studies (GWAS), researchers scan the genomes of diverse individuals to find SNP markers statistically linked to specific traits, such as disease resistance or nutritional quality (Mourad et al., 2018; Roorkiwal et al., 2022). Once identified, these markers can be used in MAS to select for desirable plants at an early stage based on their genotype rather than waiting for phenotypic expression (Collard & Mackill, 2008). The most advanced application is Genomic Selection (GS), where all SNP data of an individual is used to calculate a "Genomic Estimated Breeding Value" that predicts its overall genetic potential, dramatically accelerating the rate of crop improvement (Calus, 2010).

1.4 Genetic linkage mapping and QTL analysis

1.4.1 Genetic linkage maps: Charting the code of inheritance

A foundational tool for dissecting the genetic basis of traits is the genetic linkage map. A linkage map serves as a 'road map' of an organism's chromosomes, illustrating the relative positions of and distances between genetic markers (Collard et al., 2005; Paterson, 1996). The construction of these maps is based on a fundamental biological process: meiotic recombination. During the formation of gametes in sexually reproducing organisms, homologous chromosomes exchange segments in a process known as crossing-over. Genes and associated markers that are physically located close to one another on a chromosome tend to be inherited together as a single block; they are genetically linked (Collard et al., 2005). Conversely, the probability of a crossover event occurring between two genetic loci increases with the physical distance separating them. By analysing a segregating population, the frequency of recombination between markers can be calculated and used as a measure of genetic map distance in units called centiorgan (cM) (Collard et al., 2005).

In plant genetics, linkage maps have historically been developed using populations derived from homozygous inbred lines, such as F_2 or backcross populations (Ritter et al., 1990). In these cases, the grandparental origin of alleles and their arrangement on homologous chromosomes (the linkage phase) are known, making the identification of recombination events relatively straightforward (Jansen, 2005). However, for many species, particularly outcrossing perennials like forest trees and numerous potential industrial crops, obtaining homozygous inbred lines is difficult or impossible due to long generation times, self-incompatibility, or severe inbreeding depression (Margarido et al., 2007; Ritter et al., 1990).

Constructing linkage maps in these heterozygous, outcrossing species presents a unique set of challenges (Jansen, 2005). The parental genotypes are heterozygous, meaning linkage phase is typically unknown. Furthermore, marker loci can be polymorphic in one or both parents, leading to various segregation patterns (e.g., 1:1, 1:2:1, 3:1) within a single F_1 full-sib

family (Garcia et al., 2006; Jansen, 2005). This genetic complexity means that recombination events cannot be observed unambiguously from the raw marker data (Jansen, 2005)

To overcome these obstacles, specialized strategies have been developed. A traditional approach is the "double pseudo-testcross" method, which constructs two separate maps—one for each parent—using markers that are heterozygous in one parent but null in the other (segregating 1:1) (Garcia et al., 2006; Margarido et al., 2007). A significant limitation of this strategy is that the two parental maps can only be integrated into a single, consensus map by using "bridge" markers that are heterozygous in both parents ("intercross markers") (Garcia et al., 2006). More advanced statistical methods and software have since been developed to handle the complexities of outcrossing populations directly. These methods employ a maximum-likelihood framework to simultaneously estimate both recombination frequencies and the most probable linkage phases between all types of markers, enabling the construction of a single, integrated map from the outset (Garcia et al., 2006; Margarido et al., 2007)

1.4.2 Quantitative Trait Loci (QTL): Unravelling complex traits

Many of the most important traits in agriculture, such as crop yield, quality, and the production of valuable natural products like rubber, are quantitative traits (Collard et al., 2005). They exhibit continuous variation and are influenced by the combined action of multiple genes and their interaction with the environment (Collard et al., 2005). The specific genomic regions containing genes that affect a quantitative trait are known as QTL (Van Ooijen, 1999).

QTL mapping was at the time of study the primary methodology for identifying these genomic regions and dissecting the genetic architecture of complex traits (Collard et al., 2005; N. Lu et al., 2019). The process leverages a genetic linkage map and detailed phenotypic measurements from a mapping population to find statistical associations between genetic markers and the observed trait variation (Collard et al., 2005).

Statistical methods, such as composite interval mapping (CIM), are used to test for the presence of a QTL at various positions along the linkage groups (van Ooijen, 1999). The strength of this association is typically expressed as a logarithm of odds (LOD) score. This score represents the likelihood of a QTL being present at a specific location versus the likelihood of its absence (Risch, 1992). A graphical plot of the LOD score across a linkage group reveals peaks, with the highest peak indicating the most probable location of the QTL (Voorrips, 2002). To be declared statistically significant, this peak must exceed a predetermined threshold, which accounts for the numerous tests performed across the entire genome to minimize the rate of false positives (van Ooijen, 1999). Identifying QTL that control a complex trait like rubber content is the first critical step toward understanding its genetic control and developing strategies for its targeted improvement through MAS.

1.5 Objectives of the thesis

Building upon the critical global demand for natural rubber, the inherent vulnerabilities of *Hevea brasiliensis*, and the demonstrated potential of TKS as a temperate-climate alternative, this thesis aims to contribute to the genetic improvement and domestication of Russian dandelion for sustainable rubber production. Despite its promise, detailed understanding of the genetic architecture of rubber yield and quality traits in TKS remains limited. This study seeks to bridge this knowledge gap by employing advanced molecular breeding strategies.

Specifically, the objectives of this study are:

I) To conduct comprehensive genetic analysis of *Taraxacum* populations using AFLP markers to assess genetic diversity.

II) To construct high-resolution genetic maps of TKS, leveraging the identified molecular markers to serve as a foundational resource for future genetic studies and breeding efforts.

III) To perform QTL mapping in an outcrossing segregating population to identify genomic regions associated with natural rubber content.

IV) To elucidate the genetic architecture of this complex trait, providing insights crucial for future marker-assisted selection and genomic selection strategies in TKS breeding programs.

2 Material and Methods

2.1 Plant material

2.1.1 *Taraxacum* accessions and collections

Plant material of 96 different accessions of *Taraxacum kok-saghyz* (TKS), *T. officinale* (TO) and *T. brevicorniculatum* (TB) was obtained from the breeding partner ESKUSA GmbH and the USDA-ARS National Plant Germplasm System (**Table 2.1**).

Table 2.1: List of *Taraxacum* accessions used for AFLP analysis with respective origin and coordinates (where available) as well as providing source.

Accession name	Taxonomy	Origin	Coordinates	Source
TKS 1 to TKS 73	*T. kok-saghyz*	unknown	unknown	ESKUSA GmbH
W6 35156	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.02542 N, 79.24029 E	USDA-ARS
W6 35159	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.97208 N, 79.23467 E	USDA-ARS
W6 35160	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.94757, N 79.61324 E	USDA-ARS
W6 35164	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.81425 N, 80.01559 E	USDA-ARS
W6 35165	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.79949 N, 80.08643 E	USDA-ARS
W6 35166	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.84227 N, 80.04965 E	USDA-ARS
W6 35168	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.99447 N, 80.02118 E	USDA-ARS
W6 35169	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.99841 N, 80.00758 E	USDA-ARS
W6 35170	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.02861 N, 79.97919 E	USDA-ARS
W6 35172	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.03355 N, 79.91161 E	USDA-ARS
W6 35173	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.03700 N, 79.83145 E	USDA-ARS
W6 35176	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.06190 N, 79.87393 E	USDA-ARS
W6 35177	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.97702 N, 79.73954 E	USDA-ARS
W6 35178	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.92156 N, 79.64923 E	USDA-ARS
W6 35180	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.94876 N, 79.56365 E	USDA-ARS
W6 35181	*T. kok-saghyz*	Alma-Ata, Kazakhstan	42.98879 N, 79.59675 E	USDA-ARS
W6 35182	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.04773 N, 79.65091 E	USDA-ARS
W6 35183	*T. kok-saghyz*	Alma-Ata, Kazakhstan	43.06724 N, 79.17952 E	USDA-ARS
TO ‘3n’	*T. officinale*	Germany	unknown	ESKUSA GmbH
TO ‘RT’	*T. officinale*	Germany	unknown	ESKUSA GmbH
TO ‘2n’	*T. officinale*	Southern Germany	unknown	ESKUSA GmbH
TB ‘2’	*T. brevicornicu-latum*	unknown	unknown	ESKUSA GmbH
TB ‘8’	*T. brevicornicu-latum*	unknown	unknown	ESKUSA GmbH

For AFLP-analysis a total of 91 TKS accessions, of which 73 (TKS 1 to TKS 73) were provided in form of leaf samples by the breeding partner and 18 (W6 35156 – W6 35183) were grown

from seeds of the 2008 USDA-ARS collection, three different TO subspecies and two plants from a collection of TB were used.

2.1.2 F_1 mapping population

To generate a segregating population for rubber content, a cross was made between a high rubber producer of TKS (E30-12, ~ 7 % rubber in root dry weight) as the mother plant and a low rubber producer (E55-12, ~ 2.5 % rubber in root dry weight) as the father plant. Seeds obtained from that cross were sown on moist filter paper for germination under room temperature and transferred to trays with soil after one week, where they were grown under greenhouse conditions (21/18°C day/night, 16h light). 180 individuals of the full-sib population were transferred to individual pots and kept for leaf sampling for DNA extraction and *in vitro* cultivation.

2.2 Phenotyping

2.2.1 *In vitro* propagation of the mapping population

In order to generate enough plants/clones of each genotype for replicates at the different locations for each year, the established F_1 mapping population and the parent plants (E30-12, E55-12) were propagated *in vitro* at the facilities of the cooperation partner hortilab GmbH. For this, two to three fresh young leaf samples were cut from each plant of the mapping population; the cutting sides of the leaves were wrapped in wet tissue paper and packed in 15 mL falcon tubes for shipping. Of the 180 submitted genotype samples, 177 genotypes including the parents were successfully be propagated *in vitro*. For each location and year, 7-8 clones of each genotype were provided in 7x7 cm pots in soil in order to achieve a minimum number of five replicates per genotype for each field trial.

2.2.2 Field trials and rubber quantification

For phenotyping of rubber content in the roots of each genotype of the mapping population, field trials were conducted at the three locations, i.e. Parkstetten (SR) Bavaria, Quedlinburg (QLB) Saxony-Anhalt and Groß Lüsewitz (GL) Mecklenburg Western Pomerania in the years 2016, 2017 and 2018. The three sites represent very different environmental conditions in terms of climate and soil type across south, middle and northern Germany (**Table 2.2**). For each field trial, the *in vitro* propagated clones of the mapping population (175 individuals plus parents) were provided at the end of March of each year. The plants were sorted and kept in pots for another two weeks until planting in mid-April. Five clones of each genotype were planted in the field in a completely randomized block design, where each row (five in total) represented one block containing all genotypes of the mapping population. Randomization of the blocks was carried out with a custom R-script (R Core Team, 2024). For planting, pre-

drilled holes were prepared by hand with a dibble and plants were watered once after planting and again after another two weeks in order to achieve better rooting. For the ongoing growing season from mid-April to end of October, manual weed control and occasional watering in long periods without rain was carried out in order to make sure, enough plants of each genotype for two biological replicates for rubber quantification survived. There was no additional application of fertilizer or pesticides.

Table 2.2: Distinct environmental conditions of the three field trial locations used for phenotyping the mapping population: Parkstetten (SR), Quedlinburg (QLB), and Groß Lüsewitz (GL).

Location	Site	Climate	Soil Type	Air Temperature (Annual mean)	Precipitation (Annual mean)
Parkstetten	48°55'03.6"N 12°36'29.9"E 319 m elevation	Temperate, warm/wet	Luvisol, loess-loamy	9.2 °C	684 mm/year
Quedlinburg	51°46'22.0"N 11°08'45.9"E 140 m elevation	Temperate, mild/warm	Chernozem, loamy	9.9 °C	529 mm/year
Groß Lüsewitz	54°04'15.0"N 12°19'25.9"E 45 m elevation	Temperate, mild/wet	Pseudogley-luvisol, sandy-loamy	8.9 °C	633 mm/year

All remaining plants were harvested at the end of October after a growing period of around 180 days. After harvest, each plant was washed, patted dry and topped manually with secateurs, directly above the root crown. Fresh root and leaf material was weighed and roots were subsequently dried at 60°C in drying cabinets for 24 hours until reaching a constant dry weight. Dry weight of each root was recorded and two representative replicates of each genotype were shipped to the cooperation partner lifespin GmbH (formerly known as numares AG) in Regensburg, Germany. At lifespin GmbH, the dried roots were homogenized (finely ground) separately and natural rubber (poly(cis-1,4-isoprene)) was extracted in a toluene-based solution (containing 10 % Toluene-d8, TMS and 16 mM 2,6-dimethoxyphenol as an internal standard). After centrifugation (14,500 rpm, 10 min), the supernatant was transferred to 5-mm NMR (Nuclear Magnetic Resonance) sample tubes and analysed by ^{1}H-NMR spectroscopy using a Bruker AVANCE III 400 MHz spectrometer equipped with a 5-mm BBI probe-head. For quantification of poly(cis-1,4-isoprene), lifespin GmbH uses a fully automated NMR method with their own software and specific measurement protocols. The obtained raw data was processed by lifespin GmbH and results of rubber content (percentage of root dry weight) for two replicates of each genotype of the mapping population was obtained for each location and each year.

2.2.3 Analysis of phenotypic data

Analysis of variance (ANOVA) for rubber content was performed using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in R. The two-factorial ANOVA was

calculated by a linear mixed model, where genotype, location, year, and location x year interaction were considered as fixed effects and replicates as random effects.
Descriptive analysis of the field trial data was performed in R Studio (Posit team, 2025), using package *ggplot2* (Wickham, 2016) for creation of density plots, histograms and boxplots.
For calculating best linear unbiased estimates (BLUEs) for rubber content of each genotype, a mixed model was applied using the R package *lme4*. For this model, the genotype (individual) was considered as fixed effect, whereas environments (year*location combination), genotype and replicate by environment interactions were considered as random effects.
Variance components for calculation of broad sense heritability (H^2) were obtained from random effects model analysis in *lme4*, considering genotype, location, year, as well as genotype by environment (year, location, year x location) interactions as random effects. The H^2 for the trait rubber content was calculated using the following equation:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{GY}^2}{y} + \frac{\sigma_{GL}^2}{l} + \frac{\sigma_{GLY}^2}{yl} + \frac{\sigma_R^2}{ylr}}$$

where σ_G^2 was genotypic variance, σ_{GY}^2 genotype x year variance, σ_{GL}^2 genotype x location variance, σ_{GLY}^2 genotype x year x location variance and σ_R^2 the residual variance. The terms y, l and r indicate the number of years, locations and replicates, respectively.

2.3 DNA extraction and quantification

2.3.1 DNA extraction from plant material

All genomic DNA used in this study was extracted according to a slightly modified protocol by Stein et al. (2001).

Fresh leaf tissue of *Taraxacum* plants grown under different conditions (greenhouse, polytunnel, field) was used for DNA extraction in this study. When possible, two to three young fresh leaves (200-300 mg) from the centre of the leaf rosette or an equivalent amount of leaf tissue from older plants were cut with sterilized scissors and transferred to 15 mL falcon tubes and stored on ice for transportation. For cell disruption, the leaves of each sample were put between two rollers of a leaf extractor and a total of 1.5 mL (2 x 750 µl) pre-heated (65°C) cetyltrimethylammonium bromide (CTAB) extraction buffer (final concentration: 2 % (w/v) CTAB, 200 mM Tris/HCl pH 8.0, 20 mM EDTA pH 8.0, 1.4 M NaCl, 1 % (w/v) polyvinylpyrrolidone (K30), 1 % (v/v) b-mercaptoethanol) was pipetted on top of the rollers. The resulting homogenised leaf juice/extraction buffer mixture (1-1.2 mL) was collected in 2 mL Eppendorf tubes. The rollers of the leaf extractor were rinsed and cleaned between samples and the 2 mL tubes were stored on ice until incubation for one hour at 65°C in a water bath. After the samples had cooled down to room temperature, cold 800 µl dichloromethane-isoamylalcohol (24:1) to a

maximum volume of 2 mL was added under an extractor hood for DNA extraction. After vertically shaking the tubes for 15 minutes at 4°C (customized lab shaker), the tubes were centrifuged at 12000 rcf for 15 minutes at 4°C. After centrifugation, the supernatant of each sample was transferred to single well of a 2 mL microtiter plate (96 well plate), mixed with 12.5 mL RNAse (10 mg/mL) and incubated at 37°C for 15 minutes after sealing the plate with a cap mat. For DNA precipitation, 560 µl (0.7 volume of 800 µl) of cold (-20°C) isopropanol was added to each sample and after sealing the plate again with a cap mat, the block was manually inverted 5-6 times. For DNA pelleting, the plate was centrifuged for 20 minutes at 4°C and 2000 rcf. The supernatant was carefully removed by inverting the block and tapping the plate upside down on a paper towel to remove traces of the supernatant while keeping the pellets in the wells. This step was followed by three washing steps, first by adding 700 µl washing solution I (76 % ethanol, 200 mM sodium acetate), let it incubate for 5 minutes at room temperature and centrifuged at 2000 rcf for 10 minutes. After inverting the block again to remove the supernatant as described before, this step was repeated with 700 µl of washing solution II (76 % ethanol, 10 mM ammonium acetate) and an additional washing step with 80 % ethanol. After removing the supernatant again, the pellets were air dried for 10-30 minutes and finally resuspended in 100 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0).

The extracted genomic DNA used for general PCR methods was quantified photometrically using a nanoquant plate (16 samples per quantification) on a tecan infinite m200 or with a Qubit 2.0 fluorometer (Invitrogen) and diluted accordingly to the desired concentration with ddH_20.

2.4 Molecular marker analysis

2.4.1 Amplified Fragment Length Polymorphism (AFLP) analysis

All AFLP procedures in this study were performed according to the protocol described by Vuylsteke et al. (2007) with IRD700/800-labelled primers for detection of AFLP fragments on a LI-COR 4300 DNA analyzer. All adapters and primers for AFLP analysis were synthesized by metabion GmbH (Steinkirchen, Germany). AFLP fingerprinting was performed for the set of *Taraxacum* accessions and the parents and F_1 individuals of the mapping population.

For preparation of the AFLP templates, 250 ng of genomic DNA was used per sample for a double digestion with the enzymes *Eco*RI and *Mse*I and buffers (**Table 2.3**).

The reaction was incubated at 37°C for two hours and afterwards put into a -80°C freezer for five minutes for inactivation of enzymes.

Table 2.3: Reaction components for the double digestion of genomic DNA with *Eco*RI and *Mse*I restriction enzymes.

Components	Volume µl
ddH_2O	9.6
5 x AFLP (0.05 M Tris/HCL, 0.05 M Mg-acetate, 0.25 M K-acetate)	5.0
5 x DTT (25 mM DTT)	5.0
*Eco*RI (20 U/µL, Biolabs)	0.125
*Mse*I (10 U/µL, Biolabs)	0.25
DNA (50 ng/µL)	5
Total volume (µl)	24,975

This step was followed by the ligation of double-stranded adapters for the respective restriction sites. For preparation of the double-stranded adapters (**Table 2.4**), the respective forward and reverse sequences were hybridized by incubating them for 15 minutes at 65°C and slowly cooling them down to room temperature afterwards. The final adapters were diluted and added to a ligation mix.

Table 2.4: Nucleotide sequences for the *Eco*RI and *Mse*I adapters that are subsequently ligated to the digested DNA.

Adapter		Sequence
Eco RI	Forward adaptor	5´ CTC GTA GAC TGC GTA CC 3´
Eco RI	Reverse adaptor	3´ CTG ACG CAT GGT TAA 5´
Mse I	Forward adaptor	5´ GAC GAT GAG TCC TGA G 3´
Mse I	Reverse adaptor	3´ TA CTC AGG ACT CAT 5’

The ligation mix containing the adapters (**Table 2.5**) was then added to the restricted templates, mixed and incubated at room temperature for two hours.

Table 2.5: Components for ligation of adapters of one template.

Components	Volume µl
5 x AFLP (0.05M Tris/HCL, 0.05M Mg-acetate, 0.25M K-acetate)	5.0
5 x DTT (25 mM DTT)	5.0
5 x ATP (5 mM)	5.0
EcoR I - adapter (1 M)	5.0
Mse I - adapter (10 M)	5.0
T4- Ligase (1 U/ µL)	1.0
Total volume (µl)	26,0

The resulting template was diluted 1:10 with TE buffer (180 µl TE buffer + 20 µl ligation mix) and used as template for the pre-amplification with primers E00+1 (5´ GACTGCGTAC-CAATTC+A/C 3´) and M00+1 (5´ GATGAGTCCTGAGTAA+A/C 3´). The primers contained

complementary sequences to the ligated adapters with one selective base overhang (+1) which was either A or C, depending on the planned use of selective *Eco*RI and *Mse*I primers (see primer nomenclature Appendix I). The primers with one selective base were used in the pre-amplification of corresponding restricted fragments with the components and PCR program described in **Table 2.6** and **Table 2.7**.

Table 2.6: PCR components for pre-amplification of one sample for the AFLP marker analysis protocol.

Components	**Volume µl**
ddH_2O	30.1
5x colourless PCR buffer	5.0
25 mM $MgCl_2$	5.5
10 mM dNTPs	4.0
Primer E00+1 (0.3 µg/µL)	0.1
Primer M00+1 (0.3 µg/µL)	0.1
Taq polymerase (5 U/µL, Promega)	0.2
Ligation template 1:10 dil.	5.0
Total volume (µl)	50.0

Table 2.7: PCR program for pre-amplification for the AFLP marker analysis protocol.

Step	**Temperature**	**Time**
1.	95 °C	2 min
2.	94 °C	30 sec
3.	56 °C	1 min
4.	72 °C	1 min
5.	Go to 2; 19 times	
6.	72 °C	5 min
7.	4 °C	∞

After pre-amplification, the complete template (50 µl) was diluted 1:5 with 200 µl TE buffer. 20 µl of the 1:5 dilution was again mixed with 180 µl TE buffer for a total dilution of 1:50. This dilution was used as template for every following selective amplification.

The selective (main) amplification was performed using two differently labelled E-primers (IRD 700 and IRD 800) and one M-Primer with three additional selective nucleotides (NNN) for a reduction of amplified fragments. Selective amplification components and conditions are described in **Table 2.8** and **Table 2.9**.

Table 2.8: PCR components for selective amplification of one sample for the AFLP marker analysis protocol.

Components	**Volume µl**
ddH_2O	0.9
5x colourless PCR buffer	2.0
25 mM $MgCl_2$	1.0
10 mM dNTPs	1.0
Primer E^IRD700+NNN (5 µM))	0.4
Primer E^IRD800+NNN (5 µM)	0.4
Primer M+NNN (0.3 µg/µL)	0.2
Taq polymerase (5 U/µL, Promega)	0.1
Pre-amplification template 1:50 dil.	4.0
Total volume (µl)	10.0

Table 2.9: PCR program for selective amplification for the AFLP marker analysis protocol.

Step	**Temperature**	**Time**
1.	94 °C	30 sec
2.	65 °C	30 sec
3.	72 °C	1 min
4.	94 °C	30 sec
5.	65 °C - 0.7 °C per cycle	30 sec
6.	72 °C	1 min
7.	Go to 4. 11 times	
8.	94 °C	30 sec
9.	56 °C	30 sec
10.	72 °C	1 min
11.	Go to 8. 22 times	
12.	72 °C	5 min
13.	10 °C	∞

The amplified fragments with the respective IRD labelled primers obtained in this PCR were analysed on a Licor 4300 DNA Analyzer (LI-COR Biosciences) for electrophoresis and detection on two channels for IRDye 700 and IRDye 800 simultaneously. For this, 33 cm gels (0.25 mm spacer thickness) were cast containing 6 % Long Ranger™ polyacrylamide solution (Lonza Bioscience), 7.0 M Urea and 1.2x TBE buffer (1.34 mol/l Tris, 0.44 mol/l boric acid, 0.025 mol/l EDTA). The final gel was cast using 21 mL of the polyacrylamide solution with added 15 µl of 10 % APS (Roth) and 15 µl TEMED (Roth) for polymerization in a casting plate (at least one hour). For preparation of the electrophoresis run, the gel was mounted vertically into the Licor 4300 and a pre-run for channel calibration and temperature setting (45°C) was

started (40 mA, 40 W, 1500 V, for 10-15 minutes). Meanwhile the PCR products were prepared for the run by adding 5 µl AFLP loading buffer (0.1 M NaOH, 95 % deionized formamide, 0.05 % bromophenol blue, 0.05 % xylene xylol), denaturing for 5 minutes at 95°C and transferred immediately onto ice before loading. Sample loading was done using reusable 50/48-well (1.0 µl loading volume) or 100/96-well combs (0.4 µl loading volume). The first and last lane per gel loading was loaded with a 50-700 bp sizing standard (LI-COR) labelled with the corresponding IRDye. One gel could be used for two runs. Run parameters were as follows: 40 mA, 40 W, 1500 V, 45°C; run time 3-3.5 h).

The resulting AFLP gel images were scored to obtain binary (band presence/absence) data by scoring bands digitally in the Saga-GT software (version 3.1, LI-COR biosciences) and by visual examination. For construction of dendrograms, the SAHN clustering method UPGMA in the software NTSYSpc 2.21q was used based on a similarity matrix calculated from the AFLP marker data.

AFLP markers used for linkage map construction were scored by capillary electrophoresis as described in chapter 2.6.1 Marker scoring.

2.5 Genotyping by sequencing (GBS)

A Genotyping-by-Sequencing (GBS) protocol was utilized for double digest GBS library preparation, following the method of Meyer & Kircher (2010) with minor adjustments based on a protocol by Rohland & Reich (2012).

2.5.1 Library construction and sequencing

For library construction, genomic DNA of the parents (E30-12 and E55-12) and 180 individuals of the F_1 was extracted from leaf material as described above. Since the parents of the population were to be used as pseudo-reference, they were included in the experiment in four replications, resulting in a total of 188 samples. For accurate calibration of DNA-concentration, stock DNA of all samples was quantified twice with the Qubit BR DNA assay and diluted accordingly with TE buffer to a final concentration of 20 ng/µl per sample based on the mean of both quantifications. For better handling, the 188 samples were divided into two 96-well plates and the library construction was carried out in parallel.

Genomic DNA (200 ng) of each sample was digested overnight at 37°C with 14 units each of PstI-HF (NEB Inc.) and MspI (NEB Inc.) in 1× NEB buffer 4 and 100 ng/µL BSA, in a total volume of 20 µL. After digestion, the samples were incubated at 65°C for 20 minutes to inactivate the enzymes and then used directly for adapter ligation. Illumina sequencing libraries were prepared with slight modifications to the protocol of Meyer and Kirchner (2010). For cleanup, MagNa beads (Thermo Scientific) replaced AMPure XP beads (Beckman Coulter) to reduce

cost (Rohland and Reich, 2012). The blunt-end repair step was omitted as the P57_GBS adapter mix (P5_GBS_PstI and P7_GBS_MspI oligonucleotides) provided complementary overhangs to the restriction digest ends.

Adapter ligation and fill-in were performed as described (Meyer and Kirchner, 2010). DNA was purified, eluted in 20 µL EB (10 mM Tris-HCl, pH 8.0), and used for indexing PCR with 8 µL of eluate. The PCR reaction (50 µL) included 1× Phusion HF buffer, 200 µM each dNTP, 200 nM primers IS4_indPCR.P5 and indexing primers, and 0.02 U/µL Phusion Hot Start Flex. The PCR program involved an initial 98°C for 30 s, followed by 16 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 5 s, with a final extension at 72°C for 10 minutes. Products were purified with SPRI, eluted in 25 µL EB, and quantified using the Quant-iT PicoGreen assay (Life Technologies). Indexed samples were pooled in equimolar ratios.

Pooled DNA (500 ng) was size-fractionated using a 2 % agarose gel, and DNA (150–600 bp) was excised and purified with a MinElute Spin column (Qiagen). The GBS library was eluted in 20 µL EB and analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies) with the High Sensitivity DNA kit.

Library quantification was performed by qPCR and the concentration was determined from a standard curve. The sample was diluted to 10 nM for cluster formation on an Illumina cBot (Illumina Inc.), with 200 µL of custom sequencing primer (Read1_GBS_PstI) provided per lane. Cluster formation and 1 × 100 bp single-end sequencing were conducted on two lanes (one lane per library) of the Illumina HiSeq2000, following manufacturer protocols.

2.5.2 NGS read mapping and SNP calling

Illumina sequence data were de-multiplexed using the CASAVA pipeline 1.8 (Illumina, Inc.). The following steps—read processing, mapping, SNP calling, and filtering—were performed using a custom pipeline on the Galaxy platform hosted on JKI servers.

After a quality check of all reads using FastQC, the tool TrimGalore was used to trim off low-quality base calls from the 3' end of the reads and remove adapter sequences. For the GBS workflow without a reference genome of TKS, sequence data of both parents of the mapping population were used to create a pseudo reference genome of representative sequences with the tool CD-HIT. In the next step, trimmed reads of all genotypes were aligned to the pseudo reference genome using BWA-MEM. The resulting alignments were converted to BAM format for use in SAMtools and variant calling using BCFtools Mpileup. The resulting combined VCF file was pre-filtered for quality (QUAL < 20) and minimum read depth (DP ≥ 5) using VCFtools. The resulting VCF file was filtered according to parent genotypes (homozygous parent 1 / heterozygous parent 2, homozygous parent 2 / heterozygous parent 1) using SnpSift filter for application in JoinMap ® 5.

2.6 Construction of genetic linkage maps

Linkage maps were constructed using JoinMap ® 5 software (Van Ooijen, 2018). The dataset was entered according to the population type CP (Cross-Pollination), based on the parental genotypes and their F1 progeny. This population type generates two separate genetic maps – one for each parent – since loci can segregate from one parent, from the other, or from both parents.

2.6.1 Marker scoring

SNP markers obtained from the GBS SNP calling pipeline were filtered, transformed and scored according to JoinMap CP population segregation types: 'lm × ll' represents markers with first parent (maternal) heterozygous and second parent homozygous; 'nn × np' represents markers with the first parent homozygous and the second parent (paternal) heterozygous; 'hk × hk' represents markers with both parents heterozygous. Only markers found in all individuals of the mapping population were used for genetic mapping. Individuals which had no markers or the quality of SNPs was not sufficient were excluded from map construction.

In addition to the SNP markers obtained from GBS, a total of 26 AFLP selective primer combinations (see Appendix II) were used for AFLP marker scoring in the mapping population. In collaboration with the project partner University of Münster, the resulting AFLP fragments were separated based on size using capillary electrophoresis on an ABI (Applied Biosystems) sequencer. Presence or absence of specific peaks in the electropherograms were scores as "1" (presence) or "0" (absence) in a marker matrix for all individuals of the mapping population. AFLP markers were transcribed in accordance to the JoinMap segregation types used for SNP genotyping.

All markers were subsequently examined for significant deviations from the expected segregation ratio (1:1 for lm x ll/ nn x np and 1:2:1 for hk x hk) of individual alleles with the χ^2-test (chi-square) goodness-of-fit test integrated in the JoinMap ® 5 software and partially excluded from further analysis if necessary. Completely identical loci (similarity = 1) of each parental dataset were also excluded from further analysis.

2.6.2 Estimation of linkage groups and regression mapping

Linkage groups were estimated by applying independence LOD threshold ranges from 10.0 to 30.0 in steps of 1.0. Similarity thresholds were set to show locus pairs and individual pairs with a similarity larger than 0.950. Groups for mapping were selected from the groupings tree, by selecting eight nodes which showed a stable and high number of markers at LOD scores higher than 20. The selected groups were mapped by regression mapping with Haldane's mapping function and calculation settings adjusted to using linkages with a recombination frequency smaller than 0.40 and LOD higher than 3.00, goodness-of-fit jump threshold for removal of loci

5.00, performing a ripple after adding one locus and no third round of mapping. The resulting maps were plotted in MapChart 2.3 (Voorrips, 2002).

2.7 QTL analysis and mapping

QTL analysis was performed using MapQTL 6.0 software based on the two parental maps and BLUEs calculated from phenotypic data of 174 individuals of the mapping population (see chapter 2.3.3). QTL were detected using CIM for population type CP. After conducting 1,000 permutations, a LOD threshold of 3 was established to identify significant QTL at the 95 % confidence level. QTL intervals were defined as regions exceeding the LOD threshold of 3. Markers at or flanking the peak LOD value of a QTL were considered QTL-associated markers.

In order to assign the QTL regions and markers to the physical position of the genome assembly of TKS published by Lin et al. (2022), the respective GBS contigs containing the QTL SNP markers were blasted against a custom database based on this reference genome using the NCBI BLAST+ blastn tool in the JKI galaxy platform.

The localisation of candidate genes in relation to the QTL regions was based on the TKS genomic sequence and annotation data retrieved from the Genome Warehouse (GWH, http://bigd.big.ac.cn/gwh/), under accession number GWHBCHF00000000.

3 Results

3.1 Assessment of genetic diversity

In order to get a first understanding of genetic diversity among different TKS accessions and see their relation to other *Taraxacum* species, a total of 91 different gene bank accessions of TKS, as well as three different TO species and two samples of TB were analyzed using AFLP markers.

The selective primer combination E33M41 showed 12 clearly distinguishable polymorphic bands used for presence (1) / absence (0) scoring. Based on this marker data, a similarity matrix was calculated and used for dendrogram plotting in NTSYSpc. The resulting dendrogram is shown in **Figure 3.1**.

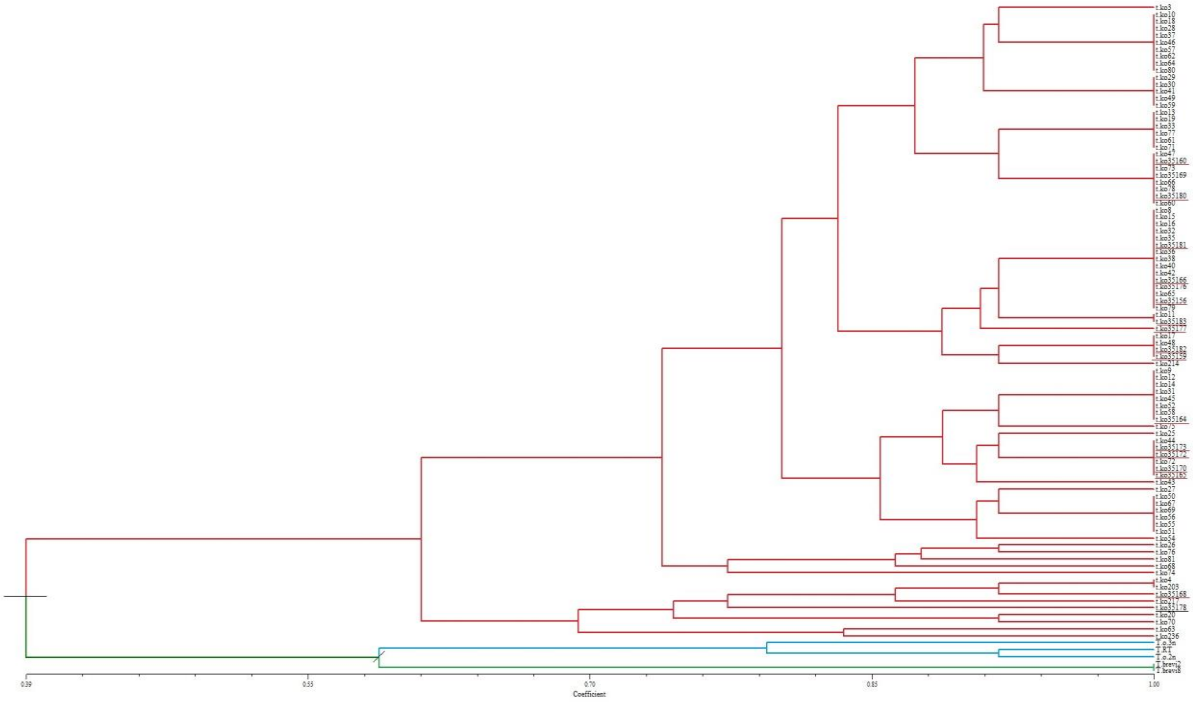

Figure 3.1: Dendrogram illustrating the genetic relationships of *T. kok-saghyz* accessions (red, TKS USDA accessions are underlined), *T. officinale* (blue) and *T. brevicorniculatum* (green).

The dendrogram confirms that the TKS (red) cluster is distinct from TO and TB, since the clusters merge at a low similarity coefficient of 0.39. While the complex internal clustering of the TKS accessions might indicate a considerable diversity within this group, the accessions of different origin (USDA and the ones provided by the breeder) show no clear distinction as they even cluster together and show very high similarity in some branches.

The TO cluster (blue) comprises three different TO species which show high similarities. While the TB cluster seems to be closer related to TO compared to TKS, it still shows a clear distinctness. The two samples of TB show very high similarity, which is not surprising due to the apomictic reproduction system of this species.

3.2 Phenotypic data of the mapping population

3.2.1 Rubber content and root weight

Out of the original 180 genotypes in the mapping population, 174, along with the parents, were successfully cloned in vitro and used in field trials at three locations over three years. In total, 3,011 root samples were successfully characterized for rubber content using NMR analysis. Rubber content showed a similar distribution across all three locations (**Figure 3.2**).

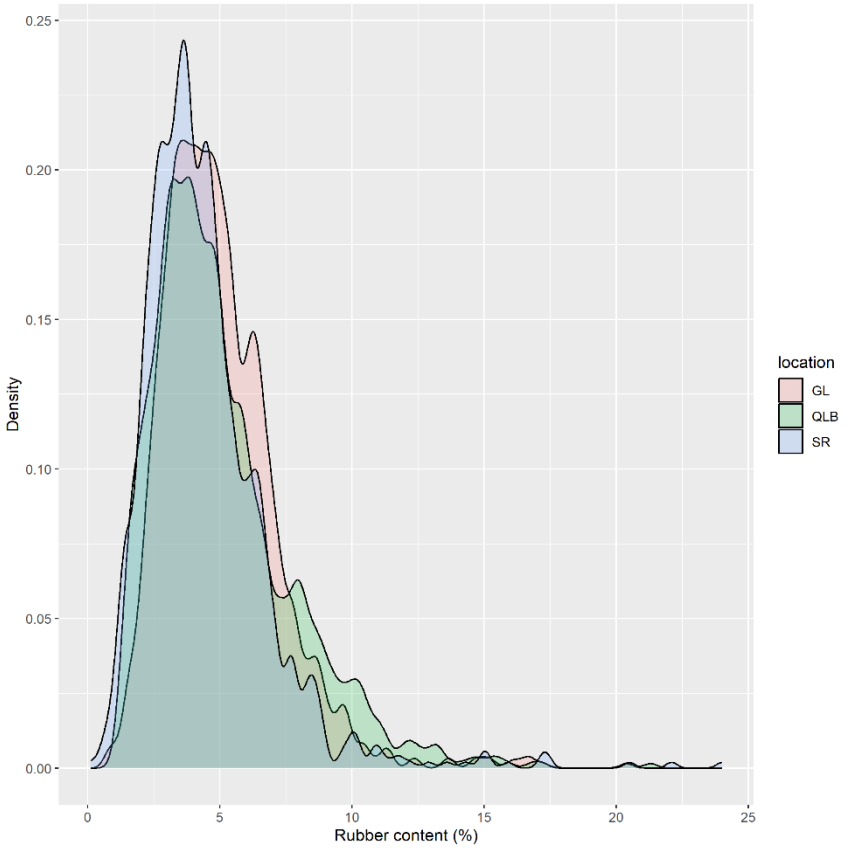

Figure 3.2: Density plot of rubber content in a mapping population of *T. kok-saghyz* (E30-12 x E55-12) over three years at the locations Groß Lüsewitz (GL), Quedlinburg (QLB) and Straubing (SR). Rubber content is shown as the percentage of root dry weight on the x-axis and the density is shown on the y-axis.

The overall results of rubber quantification demonstrate a wide range of rubber content across the samples, with a relatively low minimum value of 0.10 % and a maximal value of 24 %. The median value of 4.4 % indicates that half of the samples have a rubber content below this value, and half have a content above it. The mean value of 4.85% is slightly higher than the median, suggesting a slight positive skew in the distribution of rubber content, likely due to a few samples with very high rubber content. The overall mean of the high rubber producing parent E30-12 was 6.01 % and the mean of low rubber producing parent E55-12 was 3.72 %. The data points spread across this range highlight variability in rubber content across the studied samples. Looking into more detail for differences between locations and years (**Figure 3.3**), two-factorial ANOVA for the trait rubber content revealed highly significant ($p<0.001$) effects of

year and genotype. Location and location by year interaction also had significant (p = 0.022) effects (**Table 3.1**).

Table 3.1: Two-factorial Anlaysis of variance (ANOVA) for rubber content in the mapping population E30-12 x E55-12.

Source of variation	Sum Sq	Mean Sq	F value	p-value	
individual	282.722	1.606	4.0956	< 2e-16	***
location	2.998	1.499	3.8216	0.022	*
year	269.637	269.637	687.462	< 2e-16	***
location:year	2.994	1.497	3.8169	0.022	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*'

Broad-sense heritability (H^2) for the trait rubber content across all years and locations was estimated at 0.656.

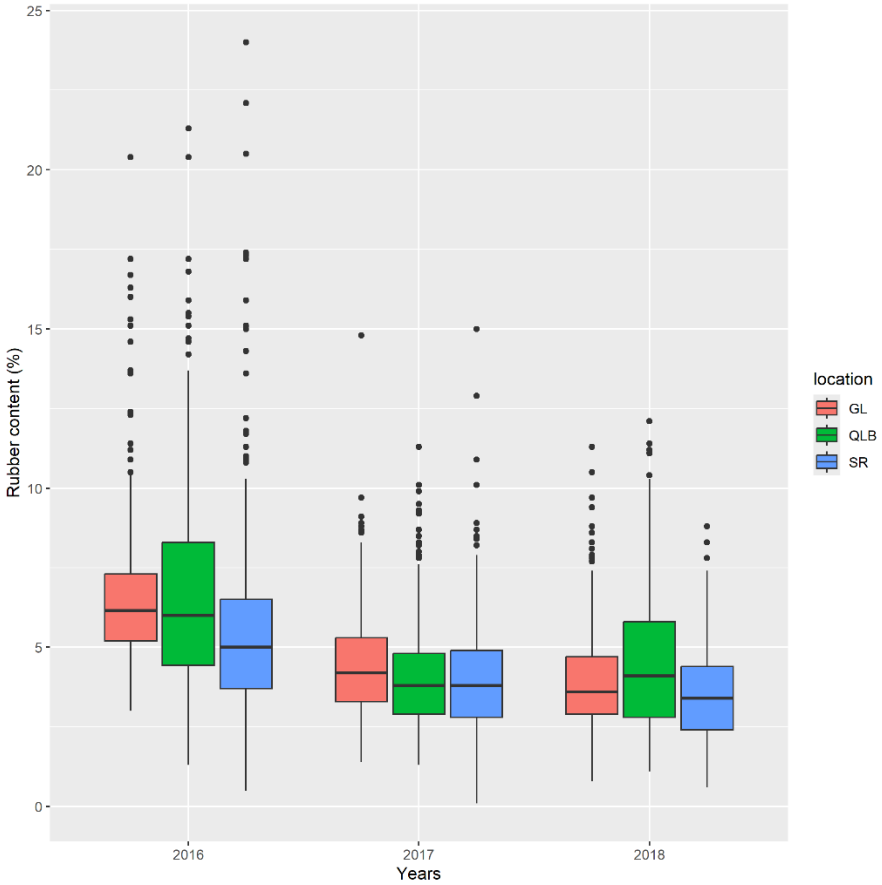

Figure 3.3: Boxplots showing the distribution of rubber content across three different locations Groß Lüsewitz (GL), Quedlinburg (QLB), Straubing (SR) over the years 2016 to 2018. Each boxplot represents the rubber content for a specific year, with the x-axis showing the years and the y-axis displaying the rubber content values in % of root dry weight. The boxes represent the interquartile range (IQR), the whiskers extend to the most extreme data points within 1.5 times the IQR, and the points outside the whiskers indicate outliers.

3.3 Genotyping-by-sequencing of the mapping population

To genotype the mapping population, two genotyping-by-sequencing (GBS) libraries were constructed, comprising 180 individuals and the parental lines. The samples were divided into two libraries for better library construction handling and higher sequencing depth on two lanes of the Illumina HiSeq 2000 platform. Each parent was represented in quadruplicate to ensure sequencing consistency and deeper coverage for pseudo-reference construction. The libraries were sequenced on the Illumina HiSeq 2000 platform, generating single-end reads with a length of 100 base pairs. A total of 62.7 GB of raw sequencing data was obtained.

Quality control analysis using MultiQC revealed high overall sequencing quality. The average guanine-cytosine (GC) content across samples was 42 %, and the mean read length post-trimming was approximately 91 bp, with a median read length of 100 bp. The average number of reads per sample was approximately 1.52 million (parents approximately 6 million reads each), providing sufficient coverage for downstream genotyping. Duplication rates were relatively high, ranging from 79.6 % to 84.8 %, which is typical for GBS libraries due to PCR amplification bias.

FastQC analysis indicated that approximately 18.2 % of reads failed quality checks, which can mainly be explained by biased sequences at the beginning of each read caused by random hexamer priming during library preparation. This is a common issue and since it does not seem to have any serious consequences for downstream analysis this was not considered a problem (Hansen et al., 2010).

3.4 Genetic maps of *Taraxacum kok-saghyz*

3.4.1 SNP calling and filtering

A comprehensive SNP discovery and filtering process was carried out to identify high-quality polymorphic markers suitable for genetic map construction. The initial SNP dataset consisted of 42,129 raw polymorphic SNPs identified across the entire mapping population following sequencing and alignment procedures. These SNPs were further filtered based on the genotype data of the parental lines, removing loci with ambiguous or non-informative parental alleles. The remaining 8,252 polymorphic SNP markers were categorized by their segregation types in Joinmap as follows:

- 2,591 markers of type <lm x ll> (heterozygous in one parent, homozygous in the other),
- 3,602 markers of type <nn x np> (the reciprocal of <lmxll>),
- 2,061 markers of type <hk x hk> (heterozygous in both parents, segregating in a 1:2:1 ratio).

In addition to SNP markers, 1071 AFLP markers resulting from AFLP fragment analysis on an ABI capillary electrophoresis were incorporated into the dataset after filtering and transposing into Joinmap format.

Of the originally 180 individuals of the mapping population, 27 individuals were excluded from the dataset, due to strict filtering allowing only 10 % missing marker data.

The final dataset used for calculating the genetic linkage maps in Joinmap 5 comprised a total of 9,323 molecular markers (8,252 SNP markers, 1,071 AFLP markers) across 153 individuals of the F_1 mapping population.

3.4.2 Construction of maternal and paternal genetic maps of TKS

Calculation of linkage groups was performed in JoinMap ® 5 using the described methods, creating maternal and paternal population nodes for further grouping. The first maternal dataset contained a total of 5,134 markers and the second paternal dataset 6,249 markers of the total 9,323 markers. Note that some of these markers were shared between the two parents, if the allele was heterozygous in both.

3.4.2.1 Maternal map of high rubber producer E30-12,

Of the 5,134 markers of the maternal dataset, 3,475 markers were excluded because they showed significant deviation from the expected segregation ratio in the χ^2-test, loci were completely identical or still had too many missing values. The remaining 1,695 loci were used in grouping for mapping. Groups were selected from the groupings tree, from nodes that showed a stable number of markers at high LOD scores (LOD = 20). This selection resulted in eight distinct linkage groups (LGs), which correspond to the basic haploid chromosome number x = 8 of the diploid TKS species. The final genetic maps calculated from these linkage groups are shown in **Figure 3.4**.

The total genetic length of the maternal map spanned 834.11 cM and was based on a total of 465 mapped markers. The overall marker density was high, with an average genetic interval of 1.83 cM between adjacent markers. The individual linkage groups exhibited variation in both length and marker density. LG3 was the longest at 132.85 cM, while LG8 was the shortest at 69.62 cM. Marker density was highest on LG1, with an average interval of just 1.35 cM, and sparsest on LG6, with an interval of 2.44 cM. The features of all maternal LGs are shown in **Table 3.2**.

Table 3.2: Summary statistics of the maternal genetic linkage map (E30-12).

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	Overall
Mapped SNP markers	86	61	68	65	49	41	42	37	449
Mapped AFLP markers	4	0	0	0	5	3	4	0	16
Average cM interval	1.35	1.69	1.98	1.44	2.04	2.44	2.33	1.93	1.83
Map length (cM)	120.50	101.08	132.85	92.05	108.16	104.97	104.89	69.62	834.11

3.4.2.2 Paternal map of low rubber producer E55-12

The paternal genetic map was constructed in the same way as parent E30-12. After excluding markers, that could distort the mapping a total of 2,115 markers were used for grouping. This map integrated a slightly larger set of polymorphic loci, with a total of 576 markers (564 SNPs and 12 AFLPs) assigned to LGs. The general characteristics of the paternal map are detailed in **Table 3.3**.

Table 3.3: Summary statistics of the paternal genetic linkage map (E55-12).

	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	Overall
Mapped SNP markers	80	67	123	93	66	64	67	4	564
Mapped AFLP markers	0	4	2	2	1	2	1	0	12
Average cM interval	2.15	1.84	2.01	2.03	1.91	2.14	1.49	13.448	2.01
Map length (cM)	161.36	124.93	195.04	210.34	125.26	143.47	99.55	40.35	1100.3

Similar to the maternal map, group selection resulted in eight linkage groups with a total of 576 mapped markers (**Figure 3.5**). The marker distribution across the paternal LGs was robust, with LG3 being the most saturated with 125 markers (123 SNPs and 2 AFLPs). The total genetic length of the paternal map was 1,100.30 cM, with an overall average marker interval of 2.01 cM. While these summary statistics suggest a well-saturated map, paternal LG8 showed significantly less mapped markers compared to the other seven LGs. It was exceptionally short, with a genetic length of only 40.35 cM, and was anchored by a mere four SNP markers. Consequently, the average marker interval on this linkage group was extremely large with 13.448 cM, indicating very poor marker coverage and resolution for this pseudo chromosome in the male parent.

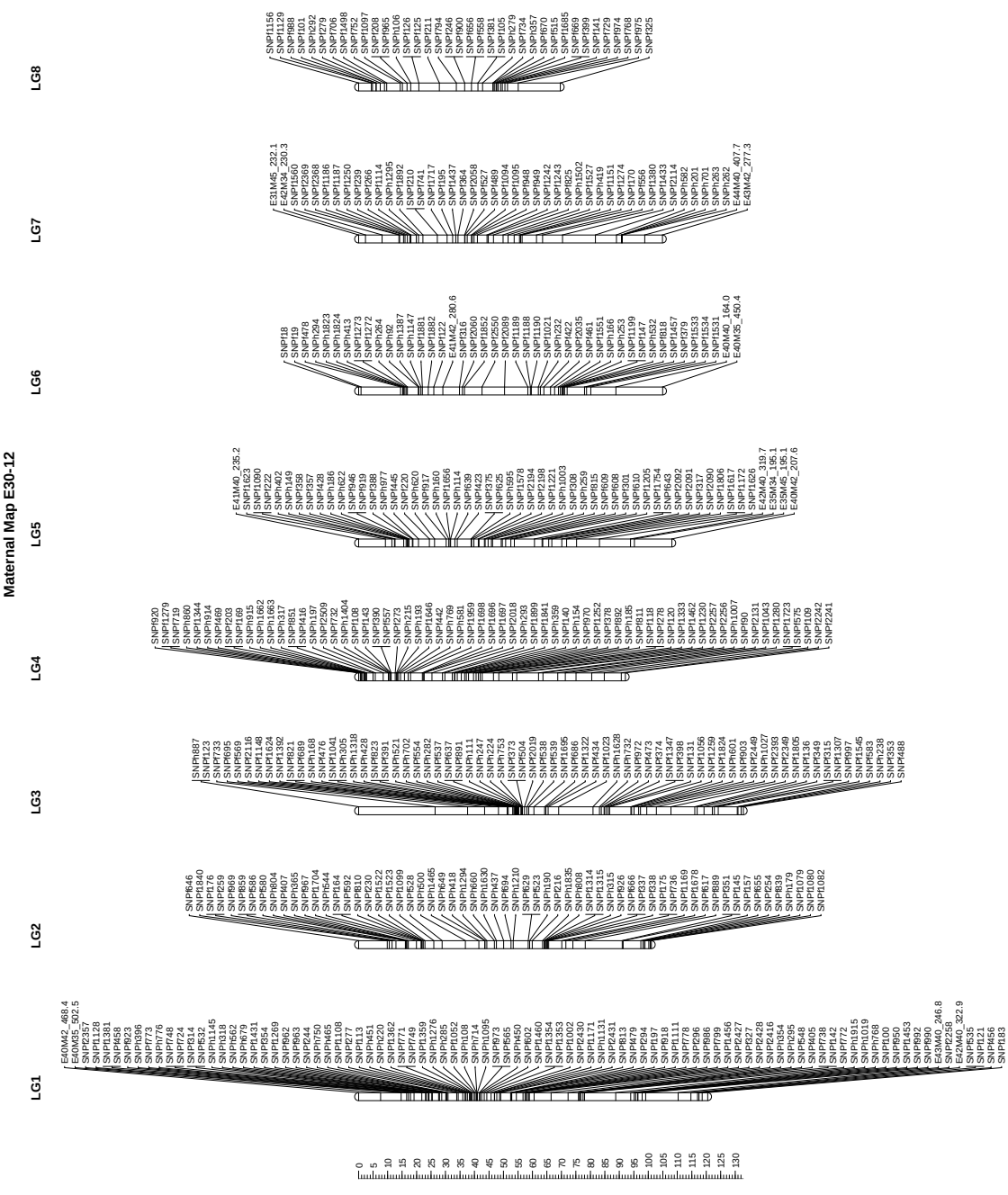

Figure 3.4: Maternal genetic linkage map of high rubber producer E30-12 based on 465 segregating molecular markers (SNP and AFLP) covering eight linkage groups.

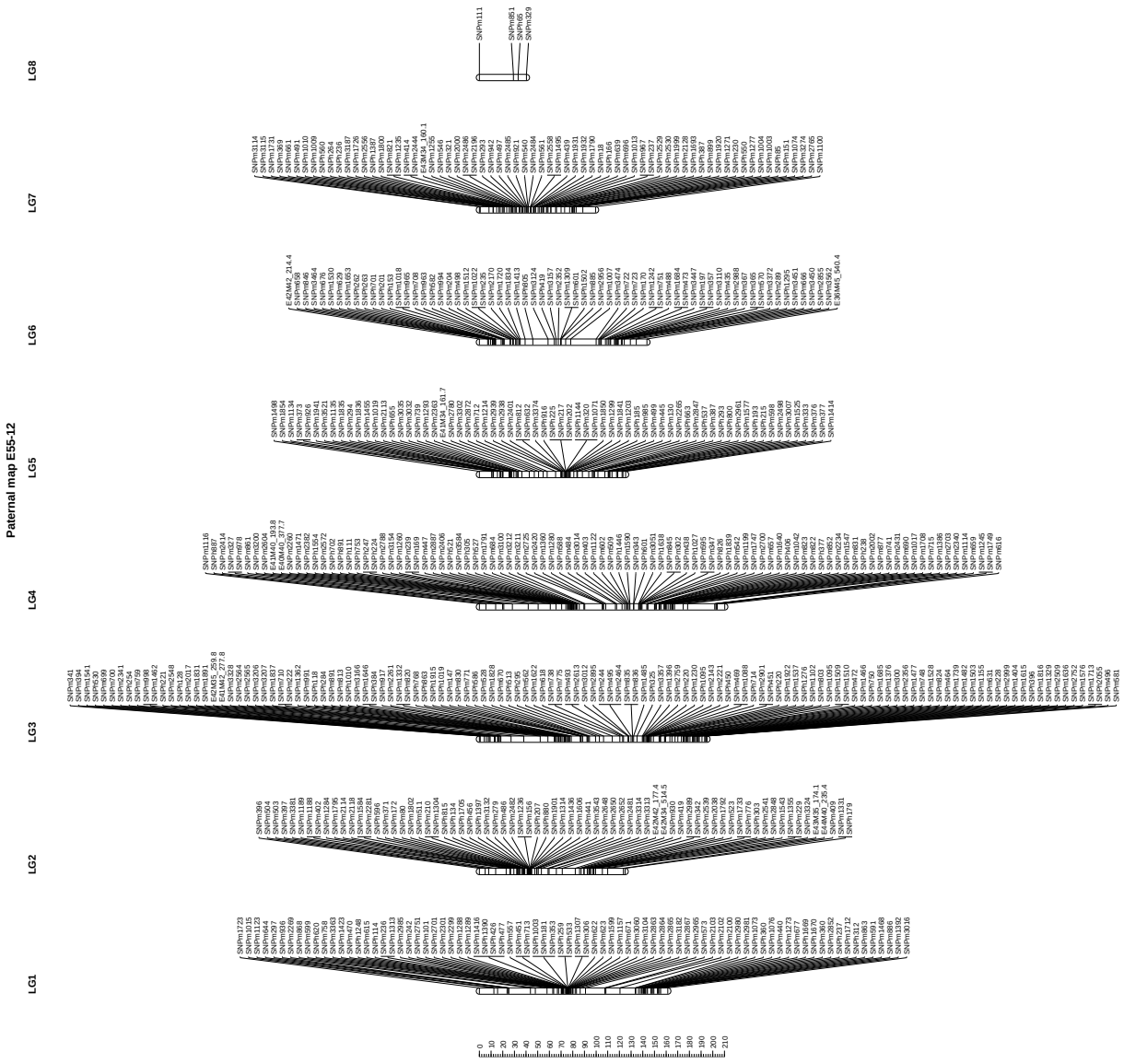

Figure 3.5: Paternal genetic linkage map of low rubber producer E55-12 based on 576 segregating molecular markers (SNP and AFLP) covering eight linkage groups.

3.5 QTL analysis for rubber content

QTL analysis was based on the collected quantitative phenotypic field data (rubber content) of the mapping population for nine environments (three locations over three years) and the calculated genetic linkage maps based on marker data derived from GBS and AFLP analysis.

The rubber content values of the final 153 individuals used for mapping, represented as BLUEs, were applied together with the marker data from the maternal and paternal maps (E30-12: 465 loci; E55: 576 loci) to calculate QTL on the eight linkage groups using the MapQTL 5 software. A total of three QTL (QTL-1, -2, and -3) with significant LOD scores > 3 were identified. Two QTL (E30-12 QTL-1, E30-12 QTL-2) are located on linkage groups 2 and 5 of the maternal genetic map (**Figure 3.6**).

Another QTL (E55-12 QTL-3) was identified on linkage group 1 of the paternal map (**Figure 3.7**).

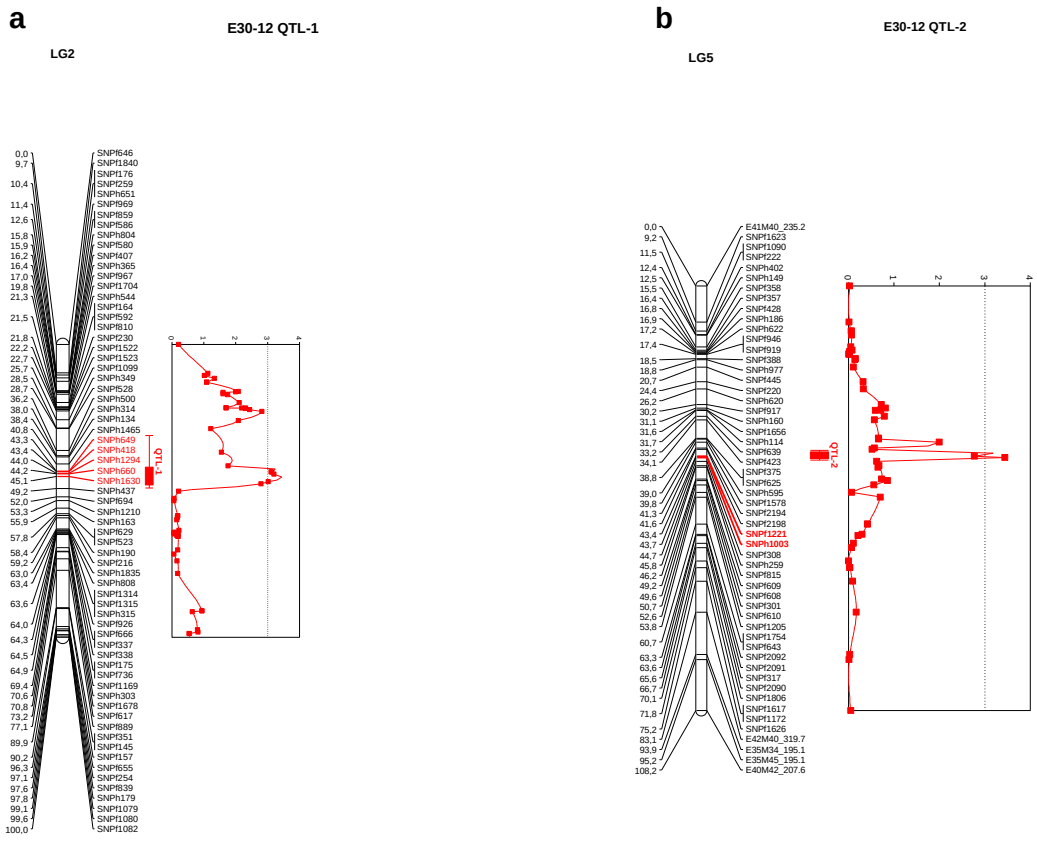

Figure 3.6: QTL analysis for rubber content on two linkage groups from the maternal map of parent E30-12. The plots display the LOD score (y-axis) at positions along the genetic map in centimorgans (x-axis). The dashed line represents the significance threshold of LOD = 3.0. **(a)** shows a significant QTL peak (E30-12 QTL-1) on linkage group 2, and **(b)** shows a second significant QTL peak (E30-12 QTL-2) on linkage group 5.

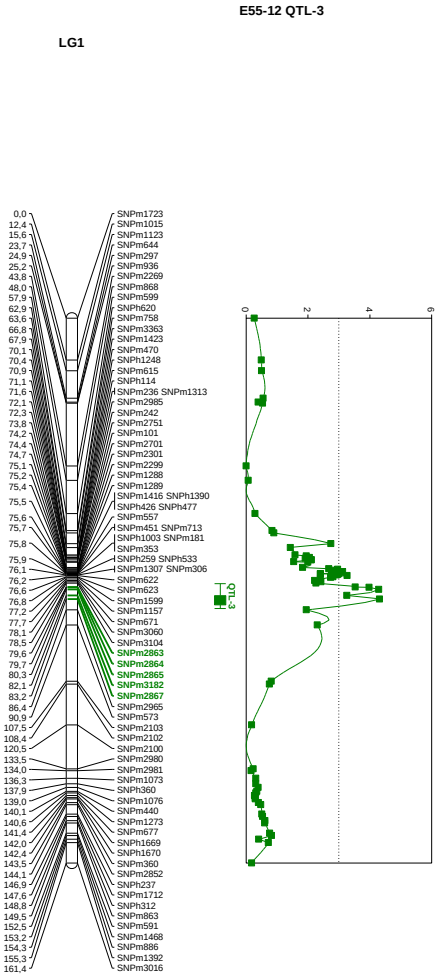

Figure 3.7: QTL analysis for rubber content showing a significant QTL peak on linkage group 1 from the paternal map of parent E55-12. The plot displays the LOD score (y-axis) at positions along the genetic map in centi-morgans (x-axis). The dashed line represents the significance threshold of LOD = 3.0.

The marker intervals of the individual QTL are given in **Table 3.4**. In total, the QTL regions comprised 12 SNP markers, each explaining 9–15 % of the phenotypic variation (PVE) in rubber content.

Table 3.4: Marker intervals for E30 QTL-1, E30 QTL-2 and E55 QTL-3 on the corresponding linkage group (LG). The position of the individual SNP markers is given in centiMorgan. All SNP markers with a LOD > 3 and the explained phenotypic variance (PVE %.) are listed. GBS contig refers to the respective contig of the pseudo-reference sequence (107 bp length) on which the respective SNP marker was called.

QTL ID	LG	Position (cM)	QTL marker	LOD	PVE %	GBS contig
E30 QTL-1	2	43.520	SNPh649	3.11	10.6	c_3913
	2	43.609	SNPh418	3.13	10.7	c_91441
	2	44.282	SNPh1294	3.20	10.8	c_42874
	2	46.853	SNPh660	3.02	12.2	c_41691
E30 QTL-2	5	43.683	SNPh1003	3.44	15.4	c_16070
E55 QTL-3	1	75.059	SNPm2299	3.12	9.1	c_55650
	1	76.200	SNPm622	3.27	9.5	c_11707
	1	79.636	SNPm2863	3.53	10.2	c_17290
	1	79.704	SNPm2864	3.98	11.4	c_17290
	1	80.320	SNPm2865	4.29	12.3	c_17290
	1	82.145	SNPm3182	3.26	9.5	c_17290
	1	83.159	SNPm2867	4.32	12.3	c_50405

The QTL regions and markers identified in this study were compared to the actual physical position in the genome assembly of TKS published by Lin et al. (2022). **Table 3.5** shows the alignment positions of the respective GBS contigs containing the QTL SNP markers on the physical chromosome maps.

Table 3.5: Blast results showing the positions of the respective QTL GBS contigs containing the significant SNP markers on the chromosomes of the reference genome.

QTL-ID	GBS contig	Chromosome (Chr)	Ident. match %	Alignment length	Chr pos. start (bp)	Chr pos. end (bp)
E30 QTL-1	c_3913	ChrA1	100	107	154133189	154133083
	c_91441	ChrA1	98.131	107	155130490	155130596
	c_42874	ChrA7	100	100	77049513	77049612
	c_41691	ChrA1	99.065	107	154779464	154779570
E30 QTL-2	c_16070	ChrA1	99.065	107	154230901	154230795
E55 QTL-3	c_55650	ChrA3	99.065	107	62843788	62843682
	c_55650	ChrA3	99.065	107	91147124	91147018
	c_55650	ChrA3	98.131	107	62838611	62838506
	c_55650	ChrA2	99.065	107	2064803	2064909
	c_55650	ChrA2	99.065	107	3068937	3069043
	c_55650	ChrA2	99.065	107	120120055	120120161
	c_55650	ChrA1	98.131	107	73264455	73264561
	c_11707	ChrA7	100	107	68854030	68853924
	c_17290	ChrA7	100	106	70103632	70103737
	c_17290	ChrA7	99.057	106	70008455	70008560
	c_50405	ChrA1	97.196	107	140245481	140245586
	c_50405	ChrA7	98	100	31180747	31180845

While the majority of GBS contigs showed unique alignments to chromosomes of the reference genome (contigs 3913, 91441, 42874, 41691, 16070, 11707), some contigs aligned to multiple positions on a chromosome (contig 17290) or even different chromosomes (contig 55650 and 50405). Contig 55650 aligned to three different chromosomes (ChrA2, ChrA3 and ChrA1) and multiple positions on chromosome ChrA2 and ChrA3. Contig 50405 aligned to ChrA1 and ChrA7. Interestingly, GBS contigs of E30 QTL-1 and E30 QTL-2 aligned mostly to ChrA1, although they were mapped on different linkage groups LG2 and LG5 of the maternal map. With exception of contig 42874, which aligned to ChrA7, the contigs of these two regions were in close proximity to each other on ChrA1, ranging between 154,133,189 – 155,130,596 bp.

Three of the four GBS contigs of E55 QTL-3 which was mapped on LG1 of the paternal map, aligned to ChrA7 (contig 11707, 17290 and 50405). As mentioned above, contig 55650 aligned to three different chromosomes and was not coherent with the other contigs of this linkage group.

Figure 3.8 shows the alignment positions of the QTL regions and their respective GBS contigs of this study in relation to the physical positions of genes involved in rubber biosynthesis and QTL identified in the Yang et al. (2023) study. Information on the rubber biosynthesis genes and their positions was also based on the Yang et al. (2023) study and annotations of the used reference genome (Lin et al., 2022). The uniquely mapped QTL contigs of this study were mainly mapped to ChrA1 and ChrA7 (indicated in red) and in close proximity to each other, indicating two major QTL regions.

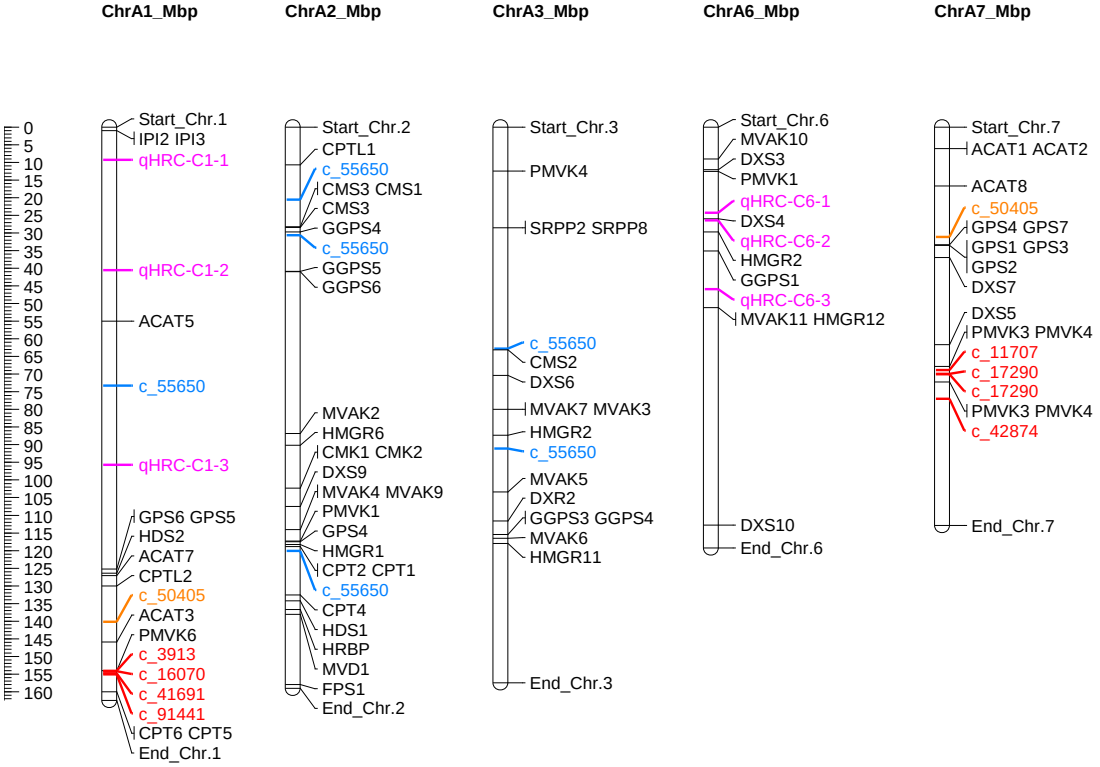

Figure 3.8: Physical maps of TKS showing the relative positions in Mbp (Mega base pairs) of genes involved in rubber biosynthesis in black colour, QTL identified by Yang et al. (2023) in pink colour, aligned GBS contigs of the QTL of this study which aligned to unique locations in red colour, contig c_55650 which aligned to three chromosomes in blue colour and contig c_50405 which aligned to ChrA1 and ChrA7 in orange colour. For MapChart data with the exact positions see Appendix III.

4 Discussion

4.1 Genetic diversity and population structure in *Taraxacum*

The assessment of genetic diversity and population structure in TKS is a critical first step for any breeding program aiming for long-term selection gains. The performed AFLP analysis provides a comprehensive view of the genetic landscape of TKS and its relationship to other *Taraxacum* species, particularly TO and TB.

The AFLP analysis and dendrogram construction demonstrated that TKS germplasm is genetically distinct from both TO and TB, with clusters merging at a low similarity coefficient of 0.39. This finding aligns with the taxonomic classification that places TKS in the section *Ceratoidea* and the common dandelion species in the section *Ruderalia* (Kirschner et al., 2015). The distinctness of these species is a prerequisite for successful interspecific hybridization, as it suggests a wide genetic distance that can be exploited to introduce new traits, such as increased biomass and root vigour from TO, into TKS breeding lines. The high similarity observed between the two TB samples is also consistent with the species' apomictic (asexual) reproductive system, which produces clonal progeny (Kirschner et al., 2013).

When examining the population structure of TKS itself, the results revealed a more complex picture. The AFLP analysis showed considerable diversity within the TKS accessions but no clear distinction between the USDA-ARS accessions from Kazakhstan and the breeder-provided accessions. This suggests that the genetic material used in breeding programs is likely derived from and closely related to the publicly available wild accessions. This conclusion is strongly supported by recent studies using SSR and SNP markers, which have also found low to moderate genetic diversity and a weak population structure in TKS. McAssey et al. (2016) reported a low average pairwise FST of 0.11 across 17 USDA-ARS accessions, which is in good accordance with the pairwise FST values ranging from 0.065 to 0.227, which were reported in the Kaiser et al. (2025) study. The conclusion that most of the genetic variation is partitioned among individuals within accessions rather than between populations (77.8% vs. 22.2%) further highlights the need to preserve diversity at the individual plant level rather than focusing solely on collecting new accessions from different geographic locations.

The lack of strong population structure in the TKS USDA-ARS accessions can be attributed to several factors. As noted by Kaiser et al. (2025), these accessions were collected from a relatively small geographic area (maximum distance of 79 km) in the Tekes River Basin of Kazakhstan, where natural barriers to gene flow are likely minimal. This is further supported by the weak and non-significant correlation found between genetic and geographical distance for the USDA-ARS accessions (Kaiser et al., 2025). This pattern is typical for a species that is an obligately outcrossing and self-incompatible, which promotes a high degree of gene flow

(Nowicki et al., 2019). This suggests that new TKS collections from different regions, particularly China, could be a valuable source of genetic variation for enriching the gene pool and preventing a genetic bottleneck in breeding programs. However, it is noteworthy that some accessions from the Zhang et al. (2021) study were found to be genetically more distinct from the USDA-ARS accessions than even the TO samples, raising questions about their proper taxonomic identification (Kaiser et al., 2025). This highlights the importance of validating taxonomic assignments, as the confusion between TKS and other *Taraxacum* species has historically been a significant issue (Kirschner et al., 2013).

In conclusion, our results, in combination with existing literature, demonstrate that the publicly available TKS germplasm possesses low to moderate genetic diversity and lacks strong population structure, which could be a limiting factor for long-term breeding progress. To overcome this, future breeding efforts must focus on: 1) the careful preservation of the genetic diversity that exists within current breeding material, 2) the integration of new, genetically diverse germplasm from other regions like China to expand the TKS gene pool, and 3) the exploitation of interspecific hybridization with other *Taraxacum* species, such as TO, to introduce traits like higher biomass and vigour.

4.2 Phenotypic variability and heritability of rubber content

The successful development of TKS as a commercial crop depends on its ability to produce high and stable yields of natural rubber. The phenotypic data from the F_1 mapping population, evaluated across three locations and three years, provide crucial insights into the genetic and environmental factors that influence this key trait. The observed extensive variability of rubber content is biologically significant as it demonstrates the existence of a broad phenotypic base for rubber production in the F_1 population, which is essential for effective selection in a breeding program (Luo et al., 2018). The positively skewed distribution of the data suggests that while most plants have a modest rubber content, a small number of individuals possess exceptionally high values, exceeding the rubber content of the high rubber producing parent, offering a clear target for selection.

The two-factorial ANOVA on rubber content revealed highly significant effects of genotype, year, and location ($p<0.001$), as well as a significant location by year interaction. The significant effect of genotype underscores the genetic basis of rubber content and confirms that selection for this trait is a viable strategy for breeders. The significant year and location effects, however, highlight the strong influence of environmental factors. The three field sites—GL, QLB, and SR—are characterized by distinct climate and soil properties. The year-to-year variation in temperature and precipitation at these sites likely contributed to the observed differences in rubber content, as previous studies have shown that factors like water supply can influence

NR content (Arias et al., 2016b). Similarly, a seasonal study showed that rubber concentration can increase during the winter months, suggesting a response to cold temperatures (Cornish et al., 2016). This interaction between genotype and environment is a key consideration for breeding, as it implies that the performance of a given genotype may vary depending on the specific growing conditions.

The finding of H^2 = 0.656 for rubber content across all years and locations is a central result of this study. This value indicates that approximately 65.6% of the total observed phenotypic variation for rubber content is attributable to genetic factors, while the remaining 34.4% is due to environmental influences and experimental error. A heritability value of this magnitude is considered moderate to high and suggests that selection for high rubber content will be highly effective, with a significant portion of the gains being heritable by the next generation. This is a very promising finding for TKS breeders. For comparison, heritability of rubber content in other alternative NR crops, such as guayule (*Parthenium argentatum*), has been estimated at a similar moderate level, with an average H^2 of 0.74 (Dierig et al., 2001). In contrast, a preliminary study on TKS by Luo et al. (2018) reported a lower $H^2$of 0.15, although their entry-mean heritability, which accounts for the effect of replication, was much higher at 0.61. This highlights the importance of multi-environment trials, as conducted in this study, for obtaining robust heritability estimates. The heritability estimate of this study, which was based on nine environments, provides a strong foundation for developing an effective selection strategy to improve TKS cultivars.

4.3 Construction of high-resolution genetic maps

4.3.1 Methodological challenges in Genotyping-by-Sequencing (GBS)

The GBS approach in this study successfully generated 42,129 raw SNPs from the TKS F_1 mapping population, which were subsequently filtered to 8,252 high-quality polymorphic markers. This substantial marker discovery demonstrates the power of GBS technology for high-throughput marker development in non-model species with limited genomic resources. However, the GBS approach in TKS revealed several important technical challenges that are characteristic of plant genomes with high repetitive content and complex genomic architectures. The primary challenge in TKS genomics is the highly repetitive structure of its genome (Lin et al., 2022). Recent genomic analyses have revealed that transposable elements (TEs), particularly Long Terminal Repeat (LTR) retrotransposons, constitute approximately 76% of the TKS genome (Lin et al., 2022). This genomic landscape poses significant challenges for next-generation sequencing approaches. Untargeted sequencing can be inefficient, and the alignment of short reads to repetitive regions is computationally complex and prone to error (Morrissey et al., 2024). The GBS protocol employed in this study was designed to circumvent these is-

sues. The selection of PstI and MspI, both methylation-sensitive enzymes, was a critical component of the experimental design (Poland et al., 2012). In plant genomes, TEs are typically silenced through epigenetic mechanisms, most notably DNA methylation, to maintain genomic integrity (P. Liu et al., 2022). Consequently, the vast repetitive fraction of the TKS genome is expected to be hypermethylated. Methylation-sensitive enzymes are blocked by CpG methylation and will not cleave DNA within these hypermethylated regions (Elshire et al., 2011). By avoiding digestion within repetitive, hypermethylated domains, the GBS library was inherently enriched for fragments from hypomethylated, low-copy, and gene-rich regions of the genome. This strategy effectively targeted sequencing efforts to the most informative and polymorphic portions of the genome, maximizing the efficiency of SNP discovery while mostly mitigating the bioinformatic challenges associated with repetitive DNA.

A common technical characteristic of GBS is the presence of high PCR duplicate levels, and the data in this study, with the value of duplication ranging from 79.6 % to 84.8 %, were in line with this expectation. While such rates are typical for GBS library preparation due to amplification bias, these PCR duplicates have relatively low (2 %) impact on the number of called genotypes, when removed beforehand (Euclide et al., 2020). Another technical issue is that the total genomic DNA isolation from plants inevitably includes a high proportion of plastid (chloroplast) DNA, which can be preferentially amplified during PCR. Because cytoplasmic genes may exhibit different mode of inheritance than nuclear genes (Cao et al., 2019), this study focussed on nuclear DNA derived SNP markers. However, this potential source of contamination was addressed by mapping all reads against available chloroplast reference sequences and keeping only the unmapped reads for further downstream analyses.

Although there are several available GBS bioinformatics pipelines like Stacks (Catchen et al., 2011), UNEAK (Lu et al., 2013) or TASSEL (Glaubitz et al., 2014), not all are suited for analysis without a reference genome or need a Unix-based computer environment (Tinker et al., 2016). Thus, this study applied a custom pipeline for SNP identification, which followed the same principles like quality trimming, building a pseudo-reference, read mapping against the pseudo-reference, variant calling, variant filtering and could be implemented in the JKI galaxy web-based platform. While difficult to compare, the number of originally 42,129 polymorphic SNPs across the mapping population and its parents seems a reasonable amount, compared to other GBS studies without a reference genome which used other pipelines (LaBonte et al., 2024).

4.3.2 Marker integration and mapping efficiency

This study successfully integrated two distinct marker technologies. While the main marker source for mapping was based on GBS-derived SNP markers, a small portion of AFLP markers could also be successfully mapped to linkage groups. However, a notable result was the low overall mapping efficiency. From an initial pool of 9,323 polymorphic markers (8,252 SNPs and

1,071 AFLPs), only 1,041 markers (465 maternal and 576 paternal) were ultimately anchored to the final maps, representing a mapping success rate of approximately 11.2 %. Such a low mapping success rate sometimes occurs in outcrossing species due to several reasons like genotyping errors due to low sequencing depth and stringent filtering of markers meeting the requirements for genetic mapping. In the study of Lu et al. (2019) for example, only 9,072 SNP markers of initially 614,295 identified polymorphic SNPs of an outcrossing population of 200 individuals of a *C. bungei* (7080) × *Catalpa duclouxii* (16-PJ-3) cross could successfully be mapped (mapping success rate of approximately 1.5 %).

The low mapping rate is not necessarily indicative of a failed experiment but rather reflects the rigorous, multi-stage quality control process that is essential for constructing accurate and reliable genetic maps, particularly in heterozygous outcrossing species. The final set of 1,041 markers of this study represents the highest-confidence subset of the initial pool, where map accuracy and marker order were prioritized over sheer marker count. The reduction in marker number occurred across several filtering stages. First, GBS is known to produce datasets with a substantial amount of missing data due to the stochastic nature of fragment sampling, especially at lower sequencing coverage (Bilton et al., 2018). The stringent requirement that markers have no more than 10 % missing data across the population already eliminated a large portion of the initial loci. Second, markers exhibiting significant segregation distortion, as determined by a χ^2 goodness-of-fit test, were removed. While segregation distortion is a common biological phenomenon in mapping populations, distorted markers can create significant artifacts, such as incorrect marker ordering and inflated map distances, and their exclusion is a standard practice to ensure map integrity (Xian-Liang et al., 2006). Finally, the linkage grouping algorithm within the JoinMap 5 ® software requires strong statistical evidence, defined by a high LOD score, to assign a marker to a linkage group. Markers with weaker or ambiguous linkage would have remained unmapped or would all be grouped in one large linkage group.

While the inclusion of AFLP markers contributed to map saturation, their dominant inheritance pattern introduces a methodological consideration. In this study, AFLP fragments were scored for presence or absence via capillary electrophoresis, treating them as dominant markers. This approach provides less genetic information than co-dominant markers, which can distinguish between homozygous and heterozygous states. An alternative, more informative approach was demonstrated by Arias et al. (2016a) in their TKS mapping study, where they used specialized software (AFLP-Quantar) to analyse band intensities and successfully score AFLP markers as co-dominant. Adopting such a method would require a different and more complex analytical pipeline but could potentially enhance the utility of AFLP markers, especially for map integration.

4.3.3 Quality and structure of the parental genetic maps

A primary achievement of this study is the successful construction of separate, high-quality genetic maps for both the maternal (high rubber producer E30-12) and paternal (low rubber producer E55-12) parents. The maternal map consists of 465 markers spanning 834.11 cM, while the paternal map contains 576 markers and covers 1100.30 cM. Crucially, both maps resolved into eight distinct linkage groups, a number that corresponds to the haploid chromosome number (n=8) of diploid TKS (Arias et al., 2016a).

The linkage analysis was performed using the CP population type in the JoinMap 5 ® software, a framework specifically designed for F_1 populations derived from heterozygous, outcrossing parents (Van Ooijen, 2018). This approach, often referred to as the "double pseudo-testcross" strategy, constructs two maps independently by analysing markers that are heterozygous in one parent and homozygous in the other, which segregate in a 1:1 ratio (Garcia et al., 2006).

An integrated or consensus map, which would merge the two parental maps into a single framework, was not constructed. This was a deliberate decision based on the available marker data. The integration of parental maps relies on a robust set of "bridge" or "intercross" markers that are heterozygous and informative in both parents, thereby serving as anchors to align homologous linkage groups (Arias et al., 2016a; Garcia et al., 2006). In this study, the primary marker types were bi-allelic SNPs and dominant AFLPs. While the initial dataset contained 2,061 SNP markers with the appropriate heterozygous segregation type (<hk x hk>), these markers were subject to the same stringent quality filtering as all others. After filtering for missing data and segregation distortion, an insufficient number of high-quality, evenly distributed bridge markers remained to confidently and accurately anchor all 16 homologous linkage groups. Forcing an integration with a sparse or unreliable anchor framework would likely have produced a chimeric and misleading map.

This highlights a key difference between marker types. Co-dominant, multi-allelic markers, such as SSR markers, are better suited for map integration. A single SSR locus can possess multiple alleles (e.g., parent 1 is genotype ab, parent 2 is cd), making it exceptionally informative for linking the two parental maps (Ahmad et al., 2018). The reliance on predominantly bi-allelic SNPs in this study, while excellent for generating high marker density, limited the potential for map integration. Thus, the presentation of two separate, high-quality parental maps represents the most accurate and robust output possible from the available data.

4.3.4 Map resolution and genome coverage

The genetic maps constructed in this study are of high resolution, representing a significant advancement for TKS genomics and providing a powerful resource for the genetic dissection of complex traits. The maternal map achieved an average marker density of one marker every 1.83 cM, while the paternal map had a density of one marker every 2.01 cM.

A comparison to previously published maps is shown in **Table 4.1**.

Table 4.1: Comparison of genetic linkage maps for TKS

	Current study (maternal)	**Current study (paternal)**	**Yang et al. (2023)**	**Arias et al. (2016a)**
Mapping population	F_1 (CP), n=153	F_1 (CP), n=153	F_1 (CP), n=127	F_1 (CP), n=94
Marker types	SNP, AFLP	SNP, AFLP	SNP (from WGR)	AFLP, SSR, COS, EST-SSR
No. of mapped Markers	465	576	12,680 bins, 322,439 SNPs	518
No. of linkage groups	8	8	9 (8 + unanchored)	8
Total genetic length (cM)	834.11	1100.30	34,220.75	894.1
Avg. marker density (cM/marker)	1.83	2.01	2.70	1.73

The map by Arias et al. (2016a), while slightly denser at 1.73 cM/marker, was constructed primarily with AFLP markers, which are generally considered less portable than SNPs and depend on available software for co-dominant scoring. Compared to the most recent high-density SNP map published by Yang et al. (2023), which reported an average marker distance of 2.70 cM, the maps from this current study demonstrate a solid resolution. Although the number of mapped markers in the Yang et al (2023) study was considerably higher, overall marker density was influenced by larger gaps, exceeding 20 cM. While this is explained by less recombination in certain chromosome segments, it may also be a result of genotyping errors due to the low coverage whole-genome resequencing (~7.4x for progeny), which consequently led to a possibly inflated genetic map length of over 34,000 cM. Inflated genetic map lengths are often artifacts of genotyping errors, which can be misinterpreted by mapping algorithms as recombination events, particularly between distant markers (Cartwright et al., 2007). In contrast, the rigorous filtering pipeline applied in the current study, while reducing the total number of mapped markers, likely produced a set of genotypes with higher confidence. Therefore, the more conservative map lengths of 834.11 cM and 1100.30 cM seem to be realistic map lengths

and comparable to other diploid plant species. Furthermore, this study had the largest population size with 153 individuals, a factor that greatly influences the accuracy of genetic maps (Ferreira et al., 2006). The combination of high marker density with a realistic map length suggests that the maps generated in this study are a robust and helpful genomic tool for future genetic analyses.

While the overall quality of both parental maps is high, the paternal map exhibits a notable anomaly in LG8. This linkage group is a significant outlier, containing only four SNP markers across a genetic distance of 40.35 cM. This results in an extremely sparse average marker interval of 13.45 cM, which contrasts sharply with the average density of 2.01 cM across the other seven paternal linkage groups. This discrepancy warrants careful consideration, as it likely indicates underlying biological or technical phenomena.

The most probable explanation for this low-resolution region is severe segregation distortion or false grouping during group selection. The methods section explicitly states that markers showing significant deviation from the expected Mendelian segregation ratios, as determined by a χ^2 test, were excluded from the mapping analysis. Segregation distortion is a well-documented biological phenomenon in plant genetics and can arise from various factors, including gamete competition, pollen-pistil incompatibility mechanisms (which are relevant in the self-incompatible TKS), or the presence of lethal or sub-lethal recessive alleles that lead to the selective loss of certain genotypes in the progeny (Xian-Liang et al., 2006). If a large chromosomal region on paternal chromosome 8 carried such a distortion-causing locus, it is highly probable that the majority of polymorphic markers within that region would have failed the χ^2 test and been filtered out prior to map construction. The four markers that were successfully mapped may reside on the flanks of this distorted region or could represent genotyping errors that, by chance, did not exhibit statistically significant distortion. The large genetic distance calculated between these few markers is likely an artifact of the mapping algorithm attempting to span a large genomic "hole" devoid of high-quality marker data, leading to an overestimation of recombination frequency.

While segregation distortion provides a strong and biologically solid hypothesis, other possibilities should be considered. First, it is conceivable that this particular chromosome in the paternal parent (E55-12) possesses a genuinely low level of heterozygosity, perhaps due to its specific ancestry or selection history (Yang et al., 2023). This would result in a true scarcity of scorable polymorphic markers in this region. Second, localized technical issues cannot be entirely excluded. The sequence of this chromosome could, for instance, be rich in motifs that interfere with the GBS protocol, such as a low density of restriction sites or the presence of sequences that inhibit PCR amplification, leading to poor data quality and marker dropout specific to this region. However, given that such technical issues are less likely to affect an entire chromosome arm uniformly, significant segregation distortion remains the most parsimonious

explanation for the observed data. This anomaly on paternal LG8 highlights the complex interplay of biological and technical factors that must be navigated in genetic mapping and underscores the importance of rigorous data filtering to produce reliable genomic resources. To further investigate this anomaly, it should be considered to use the meanwhile published genome at the chromosome level, as used in the Yang et al. (2023) study, to increase the accuracy of SNP marker positioning and solidify the arrangement of linkage groups in future.

4.4 Identification and characterization of QTL for rubber content

A primary objective of this dissertation was to elucidate the genetic architecture of natural rubber content in TKS by identifying QTL associated with this key agronomic trait. Using the genotypic data from GBS and the phenotypic BLUEs based on nine environments, QTL analysis was performed separately for the maternal (E30-12) and paternal (E55-12) genetic maps. The analysis successfully identified three significant QTL for rubber content, marking a significant step forward in understanding the genetic control of rubber production.

4.4.1 Key QTL findings and the magnitude of their effects

The QTL analysis revealed three statistically significant loci located on three different linkage groups: E30 QTL-1 on maternal LG2, E30 QTL-2 on maternal LG5, and E55 QTL-3 on paternal LG1. The significance of these QTL was robust, with peak LOD scores ranging from 3.19 to 4.32, exceeding the stringent genome-wide significance threshold of 3.0 established through a permutation test. A LOD threshold of this magnitude ensures a high degree of confidence ($p < 0.05$) that the identified QTL are not false positives (Van Ooijen, 1999).

The percentage of phenotypic variation explained (PVE) by a single QTL ranged from 9.1 % for SNPm2299 within E55 QTL-3 to a substantial 15.4 % for marker SNPh1003 within E30 QTL-2. According to the widely accepted classification, QTL that explain more than 10 % of the phenotypic variance are considered 'major' QTL (Collard et al., 2005). Therefore, E30 QTL-1 and E30 QTL2 as well as the majority of markers of E55 QTL-3 represent major genetic determinants for rubber content in this population. The identification of several major QTL, along with the likely presence of other minor-effect loci below the detection threshold of this study, supports the hypothesis that rubber content in TKS is a polygenic trait (Yang et al., 2023). This implies that while the trait is complex and influenced by many genes, these three identified loci are key contributors that can be targeted to achieve significant gains in breeding.

4.4.2 Comparison with other studies and physical localization

In order to provide context of the findings of this study, the GBS contigs associated with the identified QTL were mapped to the high-quality genome assembly of TKS published by Lin et al. (2022). This physical mapping provided critical insights when compared to other recent

studies. The markers for E30 QTL-1 (maternal LG2) and E30 QTL-2 (maternal LG5) were both physically localized to a region on chromosome A1 (ChrA1). This finding strongly corroborates the work of Yang et al. (2023), who also identified a major QTL for rubber content on ChrA1 in a different biparental population. The independent confirmation of ChrA1 as a 'hotspot' for rubber content regulation underscores its importance as a primary target for breeding efforts.

In contrast, another major QTL identified in this study, E55 QTL-3, was physically mapped to a distinct region on ChrA7. This represents a novel finding, as the study by Yang et al. (2023) did not report any significant QTL for rubber content on this chromosome. Therefore, this work not only confirms previous findings but also expands the known genetic landscape controlling rubber biosynthesis in TKS.

The process of aligning the genetic map with the physical reference genome highlighted several complexities. For instance, markers from two separate maternal linkage groups (LG2 and LG5) co-localized to ChrA1. This apparent contradiction can be attributed to the non-linear relationship between genetic and physical distance, where variable recombination rates across a chromosome can distort the genetic map relative to the physical sequence (O'Rourke, 2014; Voigt et al., 2004). Furthermore, specific mapping ambiguities were observed, such as with contig c_55650, which was associated with E55 QTL-3 but mapped to three different chromosomes (ChrA1, ChrA7, and ChrA8). Such multi-position mapping is often encountered with very short sequences. At only 107 bp, this contig may represent a repetitive element or a conserved sequence motif common across multiple gene families, making its precise physical placement ambiguous (Kerzendorfer et al., 2015). Such genomic features are common in species like TKS, which have experienced whole-genome duplication events in their evolutionary history (Lin et al., 2022).

Despite these mapping challenges, the physical locations of the QTL in relation to known rubber biosynthesis genes are highly informative. The major structural gene families, such as Small Rubber Particle Protein (SRPP), are primarily located on other chromosomes like ChrA4 (Lin et al., 2018). However, the QTL region on ChrA1 is in relative proximity (1 – 5 Mbp) to several candidate genes involved in the upstream mevalonate (MVA) pathway and precursor synthesis, including Phosphomevalonate Kinase 6 (PMVK6), cis-prenyltransferase 5 (CPT5), and CPT6. Similarly, the putatively novel QTL on ChrA7 is located on the same chromosome arm as PMVK4 and PMVK5. While the QTL peaks do not directly overlap with these genes, they are still in relatively close proximity. It strongly suggests that the identified QTL may not be allelic variants of the biosynthesis genes themselves, but rather novel regulatory loci (e.g., transcription factors) that influence the expression of these nearby candidate genes or genes involved in competing metabolic pathways (Wentzell et al., 2007). The physical distance between the QTL peak and the candidate genes underscores the importance of future fine-mapping efforts to resolve the causal gene.

4.4.3 Future directions and implications for breeding

The identification and characterization of the major QTL in this study have direct and significant implications for the genetic improvement of TKS. The SNP markers tightly linked to the QTL peaks, such as SNPm2867 within E55 QTL-3 and SNPh1003 within E30 QTL-2, are prime candidates for the development of tools for MAS. By converting these SNPs into simple, cost-effective molecular assays (e.g., KASP markers), breeders can screen large populations of seedlings for desirable alleles associated with high rubber content. This approach could dramatically increase selection efficiency and accelerate the breeding cycle by allowing for early selection without the need for laborious and time-consuming phenotyping (Collard et al., 2005).

While these QTL represent major steps forward, the identified genomic regions are still relatively large. The next logical step is to perform fine-mapping studies, similar to those conducted for agronomic traits in other crops (Lu et al., 2019), to narrow down the QTL intervals. Using larger populations and saturating the regions on ChrA1 and ChrA7 with additional markers will help pinpoint the causal genes underlying these QTL, providing deeper biological understanding and even more precise markers for breeding. Furthermore, the identification of these QTL offers new possibilities for targeted introgression (Platten & Fritsche-Neto, 2023) of the rubber content trait into high biomass TO cultivars, following the strategy of breeding interspecific hybrids for improved rubber yields (Kaiser et al., 2025).

Finally, the phenotypic data in this study revealed a significant genotype x environment interaction for rubber content. This indicates that the effect of a given genotype can vary across different environments or years. Future research should therefore focus on mapping QTL x environment interactions to identify alleles that confer not only high but also stable rubber yield across diverse conditions (Wang et al., 1999). Breeding for stability is crucial for the development of commercially viable and resilient TKS cultivars.

5 Conclusions and outlook for future research

This dissertation set out to build a foundational genomic toolkit to accelerate the genetic improvement of TKS as a sustainable, temperate-climate source of natural rubber. By integrating molecular genetics, extensive phenotyping, and quantitative trait analysis, this study successfully addressed its primary objectives and has provided critical insights and resources for the future of TKS breeding.

Beginning with an assessment of genetic diversity, AFLP marker analysis confirmed that TKS is genetically distinct from related species like TO, while also revealing that the available breeding germplasm possesses a low to moderate level of internal diversity. This highlights the importance of both preserving existing variation and strategically introducing new genetic material into breeding programs.

The subsequent multi-year, multi-location field trials of a biparental F_1 mapping population demonstrated extensive phenotypic variability for rubber content, which ranged from 0.1 % to as high as 24 %. Crucially, a broad-sense heritability of 0.656 was estimated for this trait, a significant finding that confirms rubber content is under strong genetic control and that selective breeding will be highly effective.

These phenotypic insights were then linked to the underlying genetics through the development of the first high-resolution, GBS-based genetic maps for the parental lines. The maps for the high-rubber (E30-12) and low-rubber (E55-12) parents were resolved into eight linkage groups, corresponding to the haploid chromosome number of TKS, and were saturated with 465 and 576 markers, respectively. The high-resolution genetic maps and the large SNP dataset are valuable public resources for the international TKS research community, providing a robust framework for future genetic dissection, comparative genomics and marker development.

This genomic framework enabled the final and most critical objective: the identification of QTL for rubber content. The analysis successfully pinpointed three significant, major QTL. Two were identified on the maternal map (LG2 and LG5) and one on the paternal map (LG1), with individual loci explaining between 9.1 % and 15.4 % of the phenotypic variance. By physically anchoring these QTL to the TKS reference genome, this study not only confirmed a previously reported major QTL region on chromosome A1 but also discovered a novel major QTL on chromosome A7, thereby significantly advancing our understanding of the genetic architecture of rubber production in TKS.

Detecting and characterizing large QTL for rubber content present TKS breeders with new opportunities for improvement. It takes the discipline out of tedious phenotypic selection and

introduces TKS breeders to modern, efficient, and precise breeding techniques. If SNP markers can be validated to be tightly linked to these QTL, they are practical molecular tools that can be deployed in MAS, allowing breeders to screen for high-potential individuals at the seedling stage. By confirming a key genomic region on chromosome A1 and uncovering a novel one on chromosome A7, this research provides breeders and researchers with multiple targets for validation and crop improvement. Collectively, these findings and resources could reduce key bottlenecks in the breeding cycle and significantly accelerate the journey toward developing commercially viable, high-yielding TKS cultivars or development of high biomass interspecific hybrids that can diversify and secure the global natural rubber supply.

This dissertation provides a strong foundation and simultaneously illuminates several clear paths for future research. The logical next steps should focus on translating these discoveries into practical breeding applications and deeper biological understanding.

1. **Fine-mapping and candidate gene identification:** The identified QTL regions on chromosomes A1 and A7, while significant, are still large. The immediate priority should be to fine-map these intervals using larger segregating populations, which could be based on even more diverse parents for the rubber content trait and saturating the regions with more markers. This will narrow down the location of the causal genes and facilitate the identification of candidate genes with functions related to rubber biosynthesis, precursor pathways, or metabolic regulation.
2. **Development of breeder-friendly markers for MAS:** The key SNP markers within the QTL peaks could be converted into simple, cost-effective assays (e.g., KASP markers). Validating these markers across diverse genetic backgrounds will enable their direct implementation in breeding programs for high-throughput MAS.
3. **Implementation of genomic selection (GS):** The high-density SNP data generated in this study is ideal for implementing genomic selection. GS models use genome-wide marker information to predict the overall genetic potential of an individual for a complex trait. This approach could capture the effects of numerous minor-effect QTL not detected in this study and has the potential to accelerate genetic gain even more rapidly than MAS alone.
4. **Investigating QTL stability:** The phenotypic analysis revealed a significant genotype x environment interaction. Future research should focus on evaluating the identified QTL across a wider range of environments to assess their stability. Mapping QTL x environment interactions will be crucial for identifying alleles that confer not only high but also stable rubber yields, a key requirement for any commercially successful crop.

By pursuing these avenues, the scientific community can build upon the work of this dissertation to fully unlock the genetic potential of *Taraxacum kok-saghyz* and establish it as a vital component of a more sustainable and resilient future for natural rubber.

6 Summary

The global supply of natural rubber is critically dependent on a single species, *Hevea brasiliensis*, leaving it vulnerable to disease and economic pressures. This dissertation addresses the urgent need for alternative rubber sources by focusing on the genetic improvement of Russian Dandelion (*Taraxacum kok-saghyz*, TKS), a promising temperate-climate resource crop. The primary goal was to develop genomic tools to accelerate the breeding of high-yielding TKS varieties.

To achieve this, the study first assessed the genetic landscape of TKS and related species. Using Amplified Fragment Length Polymorphism (AFLP) markers, the research confirmed that TKS is genetically distinct from related species like the common dandelion, while also revealing low to moderate genetic diversity within existing breeding germplasm. This highlights a need to strategically incorporate new genetic material to ensure long-term breeding success.

A biparental F_1 mapping population was created by crossing high and low rubber-producing parent plants. Extensive multi-year field trials across three different locations revealed significant phenotypic variation for rubber content, ranging from 0.1 % to 24 %. A high broad-sense heritability (H^2) of 0.656 was calculated for this trait, indicating strong genetic control and a high potential for improvement through selective breeding.

The core of the project involved creating the first high-resolution genetic maps for TKS using Genotyping-by-Sequencing (GBS). This resulted in two detailed parental maps. Each resolved into eight linkage groups corresponding to the haploid chromosome number of TKS. These maps served as the foundation for quantitative trait loci (QTL) analysis to identify genomic regions controlling rubber production.

The analysis successfully identified three major, statistically significant QTL associated with rubber content. Two QTL were located on the maternal map and one on the paternal map, with individual loci explaining between 9.1 % and 15.4 % of the phenotypic variance. By aligning these genetic findings with the TKS physical genome, the study confirmed a previously reported major QTL region on chromosome A1 and discovered a novel major QTL on chromosome A7.

In conclusion, this research provides a vital genomic toolkit for TKS improvement. The high-resolution maps and identified QTL for rubber content are foundational resources for modern breeding. The SNP markers linked to these QTL can now be used to develop tools for marker-assisted selection (MAS), enabling breeders to identify superior plants at an early stage, thereby accelerating the development of commercially viable TKS cultivars and helping to secure a sustainable future for natural rubber.

7 Zusammenfassung

Die globale Naturkautschuk-Versorgung ist momentan von einer einzigen Art (*Hevea brasiliensis*) abhängig, was sie anfällig für Krankheiten und wirtschaftlichen Druck macht. Diese Dissertation begegnet dem dringenden Bedarf an alternativen Kautschukquellen durch die genetische Verbesserung des Russischen Löwenzahns (*Taraxacum kok-saghyz*, TKS), einer vielversprechenden Pflanze für gemäßigte Zonen. Hauptziel war die Entwicklung genomischer Werkzeuge zur Beschleunigung der Züchtung ertragreicher TKS-Sorten.

Zunächst wurde die genetische Diversität von TKS und anderen *Taraxacum*-Arten mittels AFLP-Marker analysiert. Die Ergebnisse bestätigten die genetische Abgrenzung von TKS gegenüber verwandten Arten wie dem gewöhnlichen Löwenzahn, zeigten aber auch nur eine geringe bis mäßige genetische Vielfalt im bestehenden Zuchtmaterial auf. Dies verdeutlicht die Notwendigkeit, strategisch neues pflanzengenetisches Material zu integrieren, um einen langfristigen Zuchterfolg zu sichern.

Eine biparentale F_1-Kartierungspopulation, erzeugt aus Kreuzungen von Eltern mit hohem und niedrigem Kautschukgehalt, wurde in mehrjährigen Feldversuchen an drei Standorten untersucht. Dabei zeigte sich eine enorme phänotypische Variation des Kautschukgehalts (0,1 % bis 24 %). Die hohe Heritabilität des Merkmals (H^2 = 0,656) belegt eine starke genetische Kontrolle und signalisiert ein hohes Potenzial für die züchterische Selektion.

Im Kern dieser Arbeit stand die Erstellung der ersten hochauflösenden genetischen Karten für TKS mittels Genotyping-by-Sequencing (GBS). Es entstanden zwei detaillierte elterliche Karten mit jeweils acht Kopplungsgruppen, entsprechend der haploiden Chromosomenzahl von TKS. Diese Karten bildeten die Grundlage für die Analyse von Quantitative Trait Loci (QTL), um die für die Kautschukproduktion verantwortlichen Genomregionen zu identifizieren.

Die Analyse identifizierte drei statistisch signifikante Major-QTL für den Kautschukgehalt. Zwei davon lagen auf der mütterlichen und einer auf der väterlichen Karte. Die einzelnen Loci erklärten zwischen 9,1 % und 15,4 % der phänotypischen Varianz. Ein Abgleich mit dem physischen Genom bestätigte eine bekannte QTL-Region auf Chromosom A1 und identifizierte darüber hinaus eine bisher unbekannte QTL-Region auf Chromosom A7.

Zusammenfassend liefert diese Forschung entscheidende genomische Werkzeuge für die TKS-Züchtung. Die hochauflösenden Karten und identifizierten QTL sind eine fundamentale Ressource für die moderne Züchtung. Die mit den QTL verknüpften SNP-Marker ermöglichen die Entwicklung von Werkzeugen für die markergestützte Selektion (MAS). Damit können Züchter überlegene Pflanzen bereits im Keimlingsstadium identifizieren, was die Entwicklung kommerziell rentabler TKS-Sorten erheblich beschleunigt und zur Sicherung einer nachhaltigen Naturkautschuk-Versorgung beiträgt.

8 List of Literature

Adak, B., Chatterjee, U., & Joshi, M. (2025). Rubber-Based Sustainable Textiles and Potential Industrial Applications. *Textiles*, *5*(2), 17. https://doi.org/10.3390/textiles5020017

Ahmad, A., Wang, J.-D., Pan, Y.-B., Sharif, R., & Gao, S.-J. (2018). Development and Use of Simple Sequence Repeats (SSRs) Markers for Sugarcane Breeding and Genetic Studies. *Agronomy*, *8*(11), 260. https://doi.org/10.3390/agronomy8110260

Alekseeva, M., Rusanova, M., Rusanov, K., & Atanassov, I. (2023). A Set of Highly Polymorphic Microsatellite Markers for Genetic Diversity Studies in the Genus Origanum. *Plants*, *12*(4), 824. https://doi.org/10.3390/plants12040824

Arias, M., Hernandez, M., Remondegui, N., Huvenaars, K., Van Dijk, P., & Ritter, E. (2016a). First genetic linkage map of Taraxacum koksaghyz Rodin based on AFLP, SSR, COS and EST-SSR markers. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep31031

Arias, M., Hernández, M., & Ritter, E. (2016b). How does water supply affect Taraxacum koksaghyz Rod. Rubber, inulin and biomass production? *Industrial Crops and Products*, *91*, 310–314. https://doi.org/10.1016/j.indcrop.2016.07.024

Bae, S. W., Jung, S., Choi, S. C., Kim, M. Y., & Ryu, S. B. (2020). Lipid Composition of Latex and Rubber Particles in Hevea brasiliensis and Taraxacum kok-saghyz. *Molecules*, *25*(21), 5110. https://doi.org/10.3390/molecules25215110

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using**lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bilton, T. P., Schofield, M. R., Black, M. A., Chagné, D., Wilcox, P. L., & Dodds, K. G. (2018). *Accounting for Errors in Low Coverage High-Throughput Sequencing Data when Constructing Genetic Maps using Biparental Outcrossed Populations*. Genetics. https://doi.org/10.1101/249722

Bushman, B. S., Scholte, A. A., Cornish, K., Scott, D. J., Brichta, J. L., Vederas, J. C., Ochoa, O., Michelmore, R. W., Shintani, D. K., & Knapp, S. J. (2006). Identification and comparison of natural rubber from two Lactuca species. *Phytochemistry*, *67*(23), 2590–2596. https://doi.org/10.1016/j.phytochem.2006.09.012

Calus, M. P. L. (2010). Genomic breeding value prediction: Methods and procedures. *Animal*, *4*(2), 157–164. https://doi.org/10.1017/S1751731109991352

Cao, M., Faville, M., Jacobs, J., & Carena, M. (2019). *Genotyping-by-sequencing based cytoplasmic markers underpin population structural variation of perennial ryegrass*. Genetics. https://doi.org/10.1101/610543

Cartwright, D. A., Troggio, M., Velasco, R., & Gutin, A. (2007). Genetic Mapping in the Presence of Genotyping Errors. *Genetics*, *176*(4), 2521–2527. https://doi.org/10.1534/genetics.106.063982

Castiglioni, P., Ajmone-Marsan, P., Van Wijk, R., & Motto, M. (1999). AFLP markers in a molecular linkage map of maize: Codominant scoring and linkage group ditsribution. *Theoretical and Applied Genetics*, *99*(3–4), 425–431. https://doi.org/10.1007/s001220051253

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). *Stacks*: Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3 Genes|Genomes|Genetics*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Cherian, S., Ryu, S. B., & Cornish, K. (2019). Natural rubber biosynthesis in plants, the rubber transferase complex, and metabolic engineering progress and prospects. *Plant Biotechnology Journal*, *17*(11), 2041–2061. https://doi.org/10.1111/pbi.13181

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica*, *142*(1–2), 169–196. https://doi.org/10.1007/s10681-005-1681-5

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Cornish, K. (2001). Similarities and differences in rubber biochemistry among plant species. *Phytochemistry*, *57*(7), 1123–1134. https://doi.org/10.1016/s0031-9422(01)00097-8

Cornish, K. (2017). Alternative Natural Rubber Crops: Why Should We Care? *Technology & Innovation*, *18*(4), 244–255. https://doi.org/10.21300/18.4.2017.245

Cornish, K., Kopicky, S. L., Mcnulty, S. K., Amstutz, N., Chanon, A. M., Walker, S., Kleinhenz, M. D., Miller, A. R., & Streeter, J. G. (2016). Temporal diversity of Taraxacum koksaghyz plants reveals high rubber yield phenotypes. *Biodiversitas Journal of Biological Diversity*, *17*(2). https://doi.org/10.13057/biodiv/d170262

Dierig, D. A., Ray, D. T., Coffelt, T. A., Nakayama, F. S., Leake, G. S., & Lorenz, G. (2001). Heritability of height, width, resin, rubber, and latex in guayule (Parthenium argentatum). *Industrial Crops and Products*, *13*(3), 229–238. https://doi.org/10.1016/S0926-6690(00)00080-7

Dijk, P. J. V., Tas, I. C. Q., Falque, M., & Bakx-Schotman, T. (1999). Crosses between sexual and apomictic dandelions (Taraxacum). II. The breakdown of apomixis. *Heredity*, *83*, 715–721.

Drummond, E. B. M., & Vellend, M. (2012). Genotypic Diversity Effects on the Performance of Taraxacum officinale Populations Increase with Time and Environmental Favorability. *PLoS ONE*, *7*(2), e30314. https://doi.org/10.1371/journal.pone.0030314

Eggert, M., Schiemann, J., & Thiele, K. (2018). Yield performance of Russian dandelion transplants (Taraxacum koksaghyz L. Rodin) in flat bed and ridge cultivation with different planting densities. *European Journal of Agronomy*, *93*, 126–134. https://doi.org/10.1016/j.eja.2017.12.003

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Euclide, P. T., McKinney, G. J., Bootsma, M., Tarsa, C., Meek, M. H., & Larson, W. A. (2020). Attack of the PCR clones: Rates of clonality have little effect on RAD-seq genotype calls. *Molecular Ecology Resources*, *20*(1), 66–78. https://doi.org/10.1111/1755-0998.13087

Ferreira, A., Silva, M. F. D., Silva, L. D. C. E., & Cruz, C. D. (2006). Estimating the effects of population size and type on the accuracy of genetic maps. *Genetics and Molecular Biology*, *29*(1), 187–192. https://doi.org/10.1590/S1415-47572006000100033

Garcia, A. A. F., Kido, E. A., Meza, A. N., Souza, H. M. B., Pinto, L. R., Pastina, M. M., Leite, C. S., Silva, J. A. G. D., Ulian, E. C., Figueira, A., & Souza, A. P. (2006). Development of an integrated genetic map of a sugarcane (Saccharum spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. *Theoretical and Applied Genetics*, *112*(2), 298–314. https://doi.org/10.1007/s00122-005-0129-6

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., & Buckler, E. S. (2014). TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE*, *9*(2), e90346. https://doi.org/10.1371/journal.pone.0090346

Guyot, J., & Le Guen, V. (2018). A Review of a Century of Studies on South American Leaf Blight of the Rubber Tree. *Plant Disease*, *102*(6), 1052–1065. https://doi.org/10.1094/pdis-04-17-0592-fe

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Research*, *38*(12), e131–e131. https://doi.org/10.1093/nar/gkq224

Hasnain, H., & Mehvish, N. (2020). Assessment of plant genetic variations using molecular markers: A review. *Journal of Applied Biology & Biotechnology*. https://doi.org/10.7324/JABB.2020.80514

He, H., Wang, J., Meng, Z., Dijkwel, P. P., Du, P., Shi, S., Dong, Y., Li, H., & Xie, Q. (2024). Genome-Wide Analysis of the SRPP/REF Gene Family in Taraxacum kok-saghyz Provides Insights into Its Expression Patterns in Response to Ethylene and Methyl Jasmonate Treatments. *International Journal of Molecular Sciences*, *25*(13), 6864. https://doi.org/10.3390/ijms25136864

Hellier, B. C. (2011). Collecting in Central Asia and the Caucasus: U.S. National Plant Germplasm System Plant Explorations. *HortScience*, *46*(11), 1438–1439. https://doi.org/10.21273/HORTSCI.46.11.1438

Hodgson-Kratky, K. J. M., Stoffyn, O. M., & Wolyn, D. J. (2017). Recurrent Selection for Rubber Yield in Russian Dandelion. *Journal of the American Society for Horticultural Science*, *142*(6), 470–475. https://doi.org/10.21273/jashs04252-17

Jansen, J. (2005). Construction of Linkage Maps in Full-Sib Families of Diploid Outbreeding Species by Minimizing the Number of Recombinations in Hidden Inheritance Vectors. *Genetics*, *170*(4), 2013–2025. https://doi.org/10.1534/genetics.105.041822

Kaiser, R., Thiele, K., Stich, B., & Flüß, H. (2025). Population structure and genetic diversity of rubber-producing dandelion (Taraxacum kok-saghyz) and other Taraxacum species. *Industrial Crops and Products*, *229*, 121035. https://doi.org/10.1016/j.indcrop.2025.121035

Keeble-Gagnère, G., Pasam, R., Forrest, K. L., Wong, D., Robinson, H., Godoy, J., Rattey, A., Moody, D., Mullan, D., Walmsley, T., Daetwyler, H. D., Tibbits, J., & Hayden, M. J. (2021). Novel Design of Imputation-Enabled SNP Arrays for Breeding and Research Applications Supporting Multi-Species Hybridization. *Frontiers in Plant Science*, *12*, 756877. https://doi.org/10.3389/fpls.2021.756877

Kerzendorfer, C., Konopka, T., & Nijman, S. M. B. (2015). A thesaurus of genetic variation for interrogation of repetitive genomic regions. *Nucleic Acids Research*, *43*(10), e68–e68. https://doi.org/10.1093/nar/gkv178

Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., Heller-Uszynska, K., Jaccoud, D., Hopper, C., Aschenbrenner-Kilian, M., Evers, M., Peng, K., Cayla, C., Hok, P., & Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. In F. Pompanon & A. Bonin (Hrsg.), *Data Production and Analysis in Population Genomics* (Bd. 888, S. 67–89). Humana Press. https://doi.org/10.1007/978-1-61779-870-2_5

Kirschner, J., Štěpánek, J., Černý, T., De Heer, P., & Van Dijk, P. J. (2013). Available ex situ germplasm of the potential rubber crop Taraxacum koksaghyz belongs to a poor rubber producer, T. brevicorniculatum (Compositae–Crepidinae). *Genetic Resources and Crop Evolution*, *60*(2), 455–471. https://doi.org/10.1007/s10722-012-9848-0

Kirschner, J., Záveská Drábková, L., Štěpánek, J., & Uhlemann, I. (2015). Towards a better understanding of the Taraxacum evolution (Compositae–Cichorieae) on the basis of nrDNA of sexually reproducing species. *Plant Systematics and Evolution*, *301*(4), 1135–1156. https://doi.org/10.1007/s00606-014-1139-0

Kreuzberger, M., Hahn, T., Zibek, S., Schiemann, J., & Thiele, K. (2016). Seasonal pattern of biomass and rubber and inulin of wild Russian dandelion (Taraxacum koksaghyz L. Rodin) under experimental field conditions. *European Journal of Agronomy*, *80*, 66–77. https://doi.org/10.1016/j.eja.2016.06.011

Kuluev, B., Uteulin, K., Bari, G., Baimukhametova, E., Musin, K., & Chemeris, A. (2023). Molecular Genetic Research and Genetic Engineering of Taraxacum kok-saghyz L.E. Rodin. *Plants*, *12*(8), 1621. https://doi.org/10.3390/plants12081621

Kumar, R., Das, S. P., Choudhury, B. U., Kumar, A., Prakash, N. R., Verma, R., Chakraborti, M., Devi, A. G., Bhattacharjee, B., Das, R., Das, B., Devi, H. L., Das, B., Rawat, S., & Mishra, V. K. (2024). Advances in genomic tools for plant breeding: Harnessing DNA molecular markers, genomic selection, and genome editing. *Biological Research*, *57*(1), 80. https://doi.org/10.1186/s40659-024-00562-6

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

LaBonte, N. R., Zerpa-Catanho, D. P., Liu, S., Xiao, L., Dong, H., Clark, L. V., & Sacks, E. J. (2024). Improving precision and accuracy of genetic mapping with genotyping-by-sequencing data in outcrossing species. *GCB Bioenergy*, *16*(7), e13167. https://doi.org/10.1111/gcbb.13167

Lin, T., Xu, X., Du, H., Fan, X., Chen, Q., Hai, C., Zhou, Z., Su, X., Kou, L., Gao, Q., Deng, L., Jiang, J., You, H., Ma, Y., Cheng, Z., Wang, G., Liang, C., Zhang, G., Yu, H., & Li, J. (2022). Extensive sequence divergence between the reference genomes of Taraxacum kok-saghyz and Taraxacum mongolicum. *Science China Life Sciences*, *65*(3), 515–528. https://doi.org/10.1007/s11427-021-2033-2

Lin, T., Xu, X., Ruan, J., Liu, S., Wu, S., Shao, X., Wang, X., Gan, L., Qin, B., Yang, Y., Cheng, Z., Yang, S., Zhang, Z., Xiong, G., Huang, S., Yu, H., & Li, J. (2018). Genome analysis of Taraxacum kok-saghyz Rodin provides new insights into rubber biosynthesis. *National Science Review*, *5*(1), 78–87. https://doi.org/10.1093/nsr/nwx101

Liu, J., Fernie, A. R., & Yan, J. (2021). Crop breeding – From experience-based selection to precision design. *Journal of Plant Physiology*, *256*, 153313. https://doi.org/10.1016/j.jplph.2020.153313

Liu, P., Cuerda-Gil, D., Shahid, S., & Slotkin, R. K. (2022). The Epigenetic Control of the Transposable Element Life Cycle in Plant Genomes and Beyond. *Annual Review of Genetics*, *56*(1), 63–87. https://doi.org/10.1146/annurev-genet-072920-015534

Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S., & Costich, D. E. (2013). Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. *PLoS Genetics*, *9*(1), e1003215. https://doi.org/10.1371/journal.pgen.1003215

Lu, N., Zhang, M., Xiao, Y., Han, D., Liu, Y., Zhang, Y., Yi, F., Zhu, T., Ma, W., Fan, E., Qu, G., & Wang, J. (2019). Construction of a high-density genetic map and QTL mapping of leaf traits and plant growth in an interspecific F1 population of Catalpa bungei × Catalpa duclouxii Dode. *BMC Plant Biology*, *19*(1), 596. https://doi.org/10.1186/s12870-019-2207-y

Luo, Z., Iaffaldano, B. J., Zhuang, X., Fresnedo-Ramírez, J., & Cornish, K. (2018). Variance, Inter-Trait Correlation, Heritability, and Marker-Trait Association of Rubber Yield-Related Characteristics in Taraxacum kok-saghyz. *Plant Molecular Biology Reporter*, *36*(4), 576–587. https://doi.org/10.1007/s11105-018-1097-8

Mahmoudieh, M., Jariani, P., Dehghan, A. K., Karimi, M. R., Vadipour, F., Jahani, M., Rahman, M. M., & Naghavi, M. R. (2025). The future of rubber production: A review of genetic regulation of biosynthetic pathways and genome editing technologies in natural rubber-producing plants. *Tree Genetics & Genomes*, *21*(2). https://doi.org/10.1007/s11295-025-01690-0

Margarido, G. R. A., Souza, A. P., & Garcia, A. A. F. (2007). OneMap: Software for genetic mapping in outcrossing species: OneMap. *Hereditas*, *144*(3), 78–79. https://doi.org/10.1111/j.2007.0018-0661.02000.x

McAssey, E. V., Gudger, E. G., Zuellig, M. P., & Burke, J. M. (2016). Population Genetics of the Rubber-Producing Russian Dandelion (Taraxacum kok-saghyz). *PLOS ONE*, *11*(1), e0146417. https://doi.org/10.1371/journal.pone.0146417

Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448. https://doi.org/10.1101/pdb.prot5448

Mooibroek, H., & Cornish, K. (2000). Alternative sources of natural rubber. *Applied Microbiology and Biotechnology*, *53*(4), 355–365. https://doi.org/10.1007/s002530051627

Morgil, H., Can Gercek, Y., & Tulum, I. (2020). Single Nucleotide Polymorphisms (SNPs) in Plant Genetics and Breeding. In M. Çalışkan, O. Erol, & G. Cevahir Öz (Hrsg.), *The Recent Topics in Genetic Polymorphisms*. IntechOpen. https://doi.org/10.5772/intechopen.91886

Morrissey, A., Shi, J., James, D. Q., & Mahony, S. (2024). Accurate allocation of multimapped reads enables regulatory element analysis at repeats. *Genome Research*, *34*(6), 937–951. https://doi.org/10.1101/gr.278638.123

Mourad, A. M. I., Sallam, A., Belamkar, V., Wegulo, S., Bowden, R., Jin, Y., Mahdy, E., Bakheit, B., El-Wafaa, A. A., Poland, J., & Baenziger, P. S. (2018). Genome-Wide Association Study for Identification and Validation of Novel SNP Markers for Sr6 Stem Rust Resistance Gene in Bread Wheat. *Frontiers in Plant Science*, *9*, 380. https://doi.org/10.3389/fpls.2018.00380

Mueller, U. G., & Wolfenbarger, L. L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology & Evolution*, *14*(10), 389–394. https://doi.org/10.1016/S0169-5347(99)01659-6

Niephaus, E., Müller, B., Van Deenen, N., Lassowskat, I., Bonin, M., Finkemeier, I., Prüfer, D., & Schulze Gronover, C. (2019). Uncovering mechanisms of rubber biosynthesis in *Taraxacum koksaghyz* – role of *cis*-prenyltransferase-like 1 protein. *The Plant Journal*, *100*(3), 591–609. https://doi.org/10.1111/tpj.14471

Nowicki, M., Zhao, Y., Boggess, S. L., Fluess, H., Payá-Milans, M., Staton, M. E., Houston, L. C., Hadziabdic, D., & Trigiano, R. N. (2019). Taraxacum kok-saghyz (rubber dandelion) genomic microsatellite loci reveal modest genetic diversity and cross-amplify broadly to related species. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-38532-8

Ou, T., Wu, Z., Tian, C., Yang, Y., Gong, W., Niu, J., & Li, Z. (2025). Development of Genome-Wide SSR Markers in Leymus chinensis with Genetic Diversity Analysis and DNA Fingerprints. *International Journal of Molecular Sciences*, *26*(3), 918. https://doi.org/10.3390/ijms26030918

Ouyang, P., Kang, D., Mo, X., Tian, E., Hu, Y., & Huang, R. (2018). Development and Characterization of High-Throughput EST-Based SSR Markers for Pogostemon cablin Using Transcriptome Sequencing. *Molecules*, *23*(8), 2014. https://doi.org/10.3390/molecules23082014

Paterson, A. H. (1996). *Genome mapping in plants*. Academic Press ; R.G. Landes Co.

Peng, C., Chang, L., Wang, D., Xie, Q., Zheng, X., Xu, B., & Tong, Z. (2023). An efficient Agrobacterium tumefaciens-mediated transformation system of Taraxacum kok-saghyz Rodin (Russian dandelion). *In Vitro Cellular & Developmental Biology - Plant*, *59*(3), 413–423. https://doi.org/10.1007/s11627-023-10352-6

Platten, J. D., & Fritsche-Neto, R. (2023). Optimizing quantitative trait loci introgression in elite rice germplasms: Comparing methods and population sizes to develop new recipients via stochastic simulations. *Plant Breeding*, *142*(4), 439–448. https://doi.org/10.1111/pbr.13118

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLoS ONE*, *7*(2), e32253. https://doi.org/10.1371/journal.pone.0032253

Posit team. (2024). *RStudio: Integrated Development Environment for R* [Software]. Posit Software, PBC. http://www.posit.co/

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., & Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, *2*(3), 225–238. https://doi.org/10.1007/BF00564200

Priyadarshan, P. M. (2017). Refinements to Hevea rubber breeding. *Tree Genetics & Genomes*, *13*(1), 20. https://doi.org/10.1007/s11295-017-1101-8

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* [Software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raleira, R., & Drost, S. (2024). *Natural Rubber – Production & Consumption: Global Trade Flows and Potentially Noncompliant Indonesian Rubber Cases with the EUDR*. AidEnvironment. https://aidenvironment.org/publications/global-trade-flows-potentially-non-compliant-indonesian-rubber-cases-eudr/

Risch, N. (1992). Genetic Linkage: Interpreting Lod Scores. *Science*, *255*(5046), 803–804. https://doi.org/10.1126/science.1536004

Ritter, E., Gebhardt, C., & Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics*, *125*(3), 645–654. https://doi.org/10.1093/genetics/125.3.645

Rocha, A. C. S., Garcia, D., Uetanabaro, A. P. T., Carneiro, R. T. O., Araújo, I. S., Mattos, C. R. R., & Góes-Neto, A. (2011). Foliar endophytic fungi from Hevea brasiliensis and their antagonism on Microcyclus ulei. *Fungal Diversity*, *47*(1), 75–84. https://doi.org/10.1007/s13225-010-0044-2

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, *22*(5), 939–946. https://doi.org/10.1101/gr.128124.111

Roorkiwal, M., Bhandari, A., Barmukh, R., Bajaj, P., Valluri, V. K., Chitikineni, A., Pandey, S., Chellapilla, B., Siddique, K. H. M., & Varshney, R. K. (2022). Genome-wide association mapping of nutritional traits for designing superior chickpea varieties. *Frontiers in Plant Science*, *13*, 843911. https://doi.org/10.3389/fpls.2022.843911

Stein, N., Herren, G., & Keller, B. (2001). A new DNA extraction method for high-throughput marker analysis in a large-genome species such as Triticum aestivum. *Plant Breeding*, *120*(4), 354–356. https://doi.org/10.1046/j.1439-0523.2001.00615.x

Tinker, N. A., Bekele, W. A., & Hattori, J. (2016). Haplotag: Software for Haplotype-Based Genotyping-by-Sequencing Analysis. *G3 Genes|Genomes|Genetics*, *6*(4), 857–863. https://doi.org/10.1534/g3.115.024596

Uteulin, K., Suleimenov, B., & Pachikin, K. (2023). The Soils of Natural (In Situ) Coenopopulations of Taraxacum kok-saghyz L.E. Rodin in Kazakhstan. *Agronomy*, *13*(11), 2737. https://doi.org/10.3390/agronomy13112737

Van Ooijen, J. W. (1999). LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity*, *83*(5), 613–624. https://doi.org/10.1038/sj.hdy.6886230

Van Ooijen, J. W. (2018). *JoinMap ®5, software for the calculation of genetic linkage maps in experiment popu-lations of diploid species* [Software]. Kyazma B.V.

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, *39*(3), 312–328. https://doi.org/10.1590/1678-4685-GMB-2016-0027

Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*, *93*(1), 77–78. https://doi.org/10.1093/jhered/93.1.77

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. V. D., Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research*, *23*(21), 4407–4414. https://doi.org/10.1093/nar/23.21.4407

Voss-Fels, K., & Snowdon, R. J. (2016). Understanding and utilizing crop genome diversity via high-resolution genotyping. *Plant Biotechnology Journal*, *14*(4), 1086–1094. https://doi.org/10.1111/pbi.12456

Vuylsteke, M., Peleman, J. D., & Van Eijk, M. J. (2007). AFLP technology for DNA fingerprinting. *Nature Protocols*, *2*(6), 1387–1398. https://doi.org/10.1038/nprot.2007.175

Vuylsteke, M., Peleman, J. D., & van Eijk, M. J. T. (2007). AFLP technology for DNA fingerprinting. *Nature protocols*, *2*(6), 1387–1398. https://doi.org/10.1038/nprot.2007.175

Wang, D. L., Zhu, J., Li, Z. K. L., & Paterson, A. H. (1999). Mapping QTLs with epistatic effects and QTL×environment interactions by mixed linear model approaches: *Theoretical and Applied Genetics*, *99*(7–8), 1255–1264. https://doi.org/10.1007/s001220051331

Warren-Thomas, E., Dolman, P. M., & Edwards, D. P. (2015). Increasing Demand for Natural Rubber Necessitates a Robust Sustainability Initiative to Mitigate Impacts on Tropical Biodiversity: Rubber sustainability and biodiversity. *Conservation Letters*, *8*(4), 230–241. https://doi.org/10.1111/conl.12170

Warren-Thomas, E. M., Edwards, D. P., Bebber, D. P., Chhang, P., Diment, A. N., Evans, T. D., Lambrick, F. H., Maxwell, J. F., Nut, M., O'Kelly, H. J., Theilade, I., & Dolman, P. M. (2018). Protecting tropical forests from the rapid expansion of rubber using carbon payments. *Nature Communications*, *9*(1), 911. https://doi.org/10.1038/s41467-018-03287-9

Wentzell, A. M., Rowe, H. C., Hansen, B. G., Ticconi, C., Halkier, B. A., & Kliebenstein, D. J. (2007). Linking Metabolic QTLs with Network and cis-eQTLs Controlling Biosynthetic Pathways. *PLoS Genetics*, *3*(9), e162. https://doi.org/10.1371/journal.pgen.0030162

Whaley, W. G., & Bowen, J. S. (1947). Whaley (1947) Russian dandelion—An emergency source of natural rubber. *Washington, D.C. : United Sttes Department of Agriculture library*.

Wickham, H. (with Sievert, C.). (2016). *ggplot2: Elegant graphics for data analysis* (Second edition). Springer.

Wollenweber, T. E., Van Deenen, N., Roelfs, K.-U., Prüfer, D., & Gronover, C. S. (2021). Microscopic and Transcriptomic Analysis of Pollination Processes in Self-Incompatible Taraxacum koksaghyz. *Plants*, *10*(3), 555. https://doi.org/10.3390/plants10030555

Xian-Liang, S., Xue-Zhen, S., & Tian-Zhen, Z. (2006). Segregation distortion and its effect on genetic mapping in plants. *Chinese Journal of Agricultural Biotechnology*, *3*(3), 163–169. https://doi.org/10.1079/CJB2006110

Yan, P., Zhang, W., Hao, J., Miao, X., Wu, J., Xie, Z., Li, Z., Zhang, L., & Zhang, H. (2025). RAD-Seq-derived SSR markers: A new paradigm for genetic analysis and construction of genetically improved production populations in Pinus koraiensis. *BMC Plant Biology*, *25*(1), 238. https://doi.org/10.1186/s12870-025-06243-0

Yang, Y., Qin, B., Chen, Q., Nie, Q., Zhang, J., Zhang, L., & Liu, S. (2023). Construction of the first high-density SNP genetic map and identification of QTLs for the natural rubber content in Taraxacum kok-saghyz Rodin. *BMC Genomics*, *24*(1). https://doi.org/10.1186/s12864-022-09105-3

Yirgu, M., Kebede, M., Feyissa, T., Lakew, B., Woldeyohannes, A. B., & Fikere, M. (2023). Single nucleotide polymorphism (SNP) markers for genetic diversity and population structure study in Ethiopian barley (Hordeum vulgare L.) germplasm. *BMC Genomic Data*, *24*(1), 7. https://doi.org/10.1186/s12863-023-01109-6

Yuan, B., Ding, G., Ma, J., Wang, L., Yu, L., Ruan, X., Zhang, X., Zhang, W., Wang, X., & Xie, Q. (2020). Comparison of Morphological Characteristics and Determination of Different Patterns for Rubber Particles in Dandelion and Different Rubber Grass Varieties. *Plants*, *9*(11), 1561. https://doi.org/10.3390/plants9111561

Zhang, L., He, H., Wang, J., Du, P., Wang, L., Jiang, G., Liu, L., Yang, L., Jin, X., Li, H., & Xie, Q. (2025). Genome-Wide Analysis of the Cis-Prenyltransferase (CPT) Gene Family in Taraxacum kok-saghyz Provides Insights into Its Expression Patterns in Response to Hormonal Treatments. *Plants*, *14*(3), 386. https://doi.org/10.3390/plants14030386

Zhang, R., Zhu, A., Wang, X., Yu, J., Zhang, H., Gao, J., Cheng, Y., & Deng, X. (2010). Development of Juglans Regia SSR Markers by Data Mining of the EST Database. *Plant Molecular Biology Reporter*, *28*(4), 646–653. https://doi.org/10.1007/s11105-010-0192-2

Zhang, Y., Ren, H., Zhang, X., Wang, L., Gao, Q., Abudurezike, A., Yan, Q., Lu, Z., Wang, Y., Nie, Q., Xu, L., & Zhang, Z. (2021). Genetic diversity and evolutionary patterns of *Taraxacum kok-saghyz* Rodin. *Ecology and Evolution*, *11*(12), 7917–7926. https://doi.org/10.1002/ece3.7622

Appendix I

AFLP primer nomenclature for selective primers used in this study

		E00 + A			***M00 + A***
Selective E-primer			Selective M Primer		
E31	=	E00 + AAA	M31	=	M00 + AAA
E32	=	E00 + AAC	M32	=	M00 + AAC
E33	=	E00 + AAG	M33	=	M00 + AAG
E34	=	E00 + AAT	M34	=	M00 + AAT
E35	=	E00 + ACA	M35	=	M00 + ACA
E36	=	E00 + ACC	M36	=	M00 + ACC
E37	=	E00 + ACG	M37	=	M00 + ACG
E38	=	E00 + ACT	M38	=	M00 + ACT
E39	=	E00 + AGA	M39	=	M00 + AGA
E40	=	E00 + AGC	M40	=	M00 + AGC
E41	=	E00 + AGG	M41	=	M00 + AGG
E42	=	E00 + AGT	M42	=	M00 + AGT
E43	=	E00 + ATA	M43	=	M00 + ATA
E44	=	E00 + ATC	M44	=	M00 + ATC
E45	=	E00 + ATG	M45	=	M00 + ATG
E46	=	E00 + ATT	M46	=	M00 + ATT

Appendix II

List of AFLP selective primer combinations used for genotyping mapping population EK1412

E43M34
E40M34
E42M34
E41M34
E43M35
E40M35
E42M35
E41M35
E43M40
E40M40
E42M40
E41M40
E43M42
E40M42
E42M42
E41M42
E31M34
E35M34
E36M34
E44M34
E35M40
E36M40
E44M40
E31M45
E35M45
E36M45
E44M45

Appendix III

MapChart data of aligned QTL regions to physical positions in relation to rubber biosynthesis genes

group ChrA1	Mbp	group ChrA2	Mbp	group ChrA3	Mbp	group ChrA6	Mbp	group ChrA7	Mbp
Start Chr.1	0.001	Start Chr.2	0.001	Start Chr.3	0.001	Start Chr.6	0.001	Start Chr.7	0.001
IPI2	0.976	CPTL1	10.678	PMVK4	12.378	MVAK10	9.01167	ACAT1	6.062
IPI3	0.977	c_55650	20.648	SRPP2	28.516	DXS3	12.060867	ACAT2	6.062
qHRC-C1-1	9.315	CMS3	28.204	SRPP8	28.516	PMVK1	12.513499	ACAT8	16.682
qHRC-C1-2	40.552	CMS1	28.213	c_55650	62.839	qHRC-C6-1	24.318536	c_50405	31.181
ACAT5	55.039	CMS3	28.411	CMS2	63.108	DXS4	25.962445	GPS4	33.341
c_55650	73.264	GGPS4	29.669	DXS6	70.441	qHRC-C6-2	26.486808	GPS7	33.344
qHRC-C1-3	95.730	c_55650	30.689	MVAK7	79.987	HMGR2	29.678002	GPS1	33.461
GPS6	125.211	GGPS5	40.927	MVAK3	79.988	GGPS1	35.10521	GPS3	33.461
GPS5	125.212	GGPS6	40.961	HMGR2	87.269	qHRC-C6-3	46.002048	GPS2	33.483
HDS2	126.423	MVAK2	86.810	c_55650	91.147	MVAK11	51.204785	DXS7	36.867
ACAT7	127.108	HMGR6	90.209	MVAK5	103.397	HMGR12	51.206525	DXS5	61.623
CPTL2	129.992	CMK1	102.256	DXR2	111.592	DXS10	112.787997	PMVK3	67.759
c_50405	140.245	CMK2	102.258	GGPS3	115.392	End Chr.6	119.3	PMVK4	67.759
ACAT3	145.887	DXS9	107.484	GGPS4	115.396			c_11707	68.854
PMVK6	153.956	MVAK4	114.108	MVAK6	116.454			c_17290	70.008
c_3913	154.133	MVAK9	114.116	HMGR11	117.973			c_17290	70.104
c_16070	154.231	PMVK1	117.323	End Chr.3	157.500			PMVK3	72.193
c_41691	154.779	GPS4	117.547					PMVK4	72.193
c_91441	155.130	HMGR1	118.273					c_42874	77.050
CPT6	160.007	CPT2	118.882					End Chr.7	112.900
CPT5	160.009	CPT1	118.922						
End Chr.1	162.500	c_55650	120.120						
		CPT4	132.558						
		HDS1	134.339						
		HRBP	136.732						
		MVD1	138.139						
		FPS1	157.991						
		End Chr.2	159.000						

Declaration

I declare: this dissertation submitted is a work of my own, written without any illegitimate help by any third party and only with materials indicated in the dissertation. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me. At any time during the investigations carried out by me and described in the dissertation, I followed the principles of good scientific practice as defined in the Statutes of the Justus Liebig University Giessen for the Safeguarding of Good Scientific Practice.

Place, Date/Ort, Datum Signature/ Unterschrift


JKI
Julius Kühn-Institut
Bundesforschungsinstitut für Kulturpflanzen